

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using bw model

Run on: January 2, 2003, 15:17:10 ; Search time 16.9825 Seconds
(without alignments)
2751.141 Million cell updates/sec

Title: US-09-825-414-7

Perfect score: 2412
Sequence: 1 MHINRRVQPPVATDSFRT.....IEEGTAASPSSEIPFRMRS 486

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	155.5	6.4	1122	2 G64887	probable tail fibre
2	150.5	6.2	2055	2 T31110	extracellular matr
3	147.5	6.1	1822	2 S33441	EF protein - Strept
4	147	6.1	1156	2 T34852	probable secreted
5	145.5	6.0	971	2 B90835	probable tail fibre
6	145.5	6.0	973	2 C85693	probable membrane
7	144.5	6.0	2232	2 T34434	hypothetical prote
8	138.5	5.7	2155	2 C97523	hypothetical prote
9	138.5	5.7	2155	2 AD2742	conserved hypochet
10	136	5.6	597	2 T35746	hypothetical prote
11	136	5.6	4776	2 E95206	cell wall surface
12	135.5	5.6	1731	2 B98241	hypothetical prote
13	135.5	5.6	1731	2 AB3045	ice nucleation pro
14	135.5	5.6	2479	2 F87386	conserved hypochet
15	135	5.6	718	2 A81122	hypothetical prote
16	135	5.6	1128	1 T08322	plasmid replicatio
17	134	5.6	2271	2 F90073	hypothetical prote
18	133	5.5	1208	2 C82779	hemolysin-type cal
19	132	5.5	1763	2 T17465	rifamycin polyketi
20	132	5.5	1787	2 AG1360	probable cape-meas
21	131.5	5.5	3591	1 S21010	filamentous hemaggl
22	131	5.4	556	1 A28185	formate-tetrahydro
23	130	5.4	810	2 B84185	cytochrome-like pr
24	130	5.4	989	2 AE3045	ice nucleation pro
25	130	5.4	1009	2 G98240	hypothetical prote
26	129.5	5.4	703	2 H87360	hypothetical prote
27	129.5	5.4	1541	2 T02831	AAA protein L4171.
28	129	5.3	732	2 B85518	hypothetical prote
29	129	5.3	732	2 B90668	hypothetical prote

30	129	5.3	6713	2 B89921	hypothetical prote
31	128.5	5.3	1026	2 A48995	paracrystalline su
32	128.5	5.3	1073	2 C87393	S-layer protein Ra
33	128.5	5.3	7576	2 T17428	FK506 polyketide s
34	127	5.3	1238	2 T03465	probable exonuclea
35	127	5.3	1326	2 AC3372	kinesin-like prote
36	126.5	5.2	397	2 I39579	necB protein - Alc
37	126.5	5.2	888	2 D87433	conserved hypochet
38	126.5	5.2	1109	2 A56143	surface-array prot
39	126.5	5.2	2535	2 AC0304	probable hemolysin
40	126	5.2	654	2 C87587	hypothetical prote
41	126	5.2	1147	2 T35781	hypothetical prote
42	125.5	5.2	493	2 S07375	flagellin H-1 - Sa
43	125.5	5.2	653	2 T34976	probable integral
44	125.5	5.2	1065	2 S19482	hypothetical prote
45	125.5	5.2	1430	2 AF0351	probable autocrans

ALIGNMENTS

RESULT 1

probable tail fiber protein GP37 - Escherichia coli (strain K-12)
C/Species: Escherichia coli
C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #ext_change 01-Mar-2002
C/Accession: G64887; T09189
R/Blatner, P.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: G64887
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-1122 <BLAT>
A/Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AACT4454.1; PID:g1787636.
A/Experimental source: Strain K-12, substrain MG1655
R/Alba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.; Kasai, H.
.; Motomura, K.; Nakade, S.; Nakamura, Y.; Nishimoto, H.; Nishio, Y.; Oshima, T.; Saito,
moto, Y.; Horikuchi, T.
DNA Res. 3, 363-377, 1996
A/Title: A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the
A/Reference number: Z16603; MUID:97251357; PMID:9097039
A/Accession: T09189
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 3-1122 <AIB>
A/Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AACT4454.1; PID:g1787636.

Query Match 6.4%; Score 155.5; DB 2; Length 1122;
Best Local Similarity 23.7%; Pred. No. 0.097;
Matches 121; Conservative 60; Mismatches 220; Indels 109; Gaps 21;

QY	19	RTSDASLASSSVRSVSDQREINAIADYLTGHPFAHGLPADSDAGQAVDVHNAQI 78
DB	117	RMSAVAVQNTAAKKSASDSTAREATTAAD---AASBARAASVSAGAAASAGASS 173
QY	79	TALIEFRASRLHEGETPATIAD-TPAKAKLDRLATTTSGALRATPFAPFAMSLQYQPA 137
DB	174	SA-----GTASTKATEASKSAAASSSKSAATSGAATSEBTNASLQSAATS 223
QY	138	INKGDMIPALPKLTPISGALSGAMDQVCTKMDAATGCHLHLSASPDRLHMAA--S 195
DB	224	ASTA-----TTWASEAATSARDAAASKAASGS-ETNASSSSSASASATAAGNS 272
QY	196	VK-----RHSPSLARQVLDTGVAVQF--YSARNAVRTVAPALASHPAQAQAVDLG 244
DB	273	AKAAKSETNARSSETTAAGSASAAAGSKTAAASASAAATSGAQSASATAAGKAMESA 332
QY	245	VSNAGGLAANAGFGNRLLSVQSRDHQRCALVLGLDKKEPKAQLSEENDWLEAYKAISKA 304
DB	333	ASSASTATTKAGATGQAASAAARS-----ASAAKTSSTNAAKASSETS--AESKTAAS 383

Qy 305 SYSGAALNAGKRMAGLPLDMATDAMGAVRSLSVSSLTQNLGLAGGAGVGVKGLQEMATK 364
Db 384 SASSAASSASS--ASASKDEATQSAAS--SATTASTKATEAGS----- 426
Qy 365 NITDPATKAASVOLTNLGSAVAVAGWTTAALTDPVAKKAESF-----IOD--TVK-- 414
Db 427 -----ATAAQSST--AESATRA-----ETAAKRAEDTASAVALEDASTTKGI 470
Qy 415 ---STASSTGVADQTVKLAKTV-----KDMGGEAITHGTASLRNTVNNLR----- 458
Db 471 VOLSSATNSTSETLAATPKAVKSAVDNAEKRLQKQDQNGADIPDKGCLP--NNINAVSKTDF 529
Qy 459 --QRPAREADIEGGTAASPEIPRPWRS 486
Db 530 ADKRGMYRVVNAPAGATSGYYPVVMRS 559
RESULT 2
T31110
Extracellular matrix binding protein - Abiotrophia defectiva (fragment)
C;Species: Abiotrophia defectiva
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C;Accession: T31110
R;Manganello, R.; van de Rijn, I.
Infect. Immun. 67, 50-56, 1999
A;Title: Cloning and characterization of emb, a gene encoding the major adhesin of Streptococcus pneumoniae
A;Reference number: Z20988; MUID:99081722; PMID:9864195
A;Accession: T31110
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2055 <MAN>
A;Cross-references: EMBL:AF067776; NID:g3249002; PID:g3249003; PIDN:AA003320.1
C;Genetics:
A;Gene: emb

Query Match 6.2%; Score 150.5; DB 2; Length 2055;
Best Local Similarity 22.2%; Pred. No. 0.44; Mismatches 73; Indels 123; Gaps 21;
Matches 126; Conservative 73;
Qy 3 INRRVQCPVPTA-----TDGFRITASDASLA-----SSSVRSVSSDQREINAIADYLTDH 52
Db 726 INEISQRPDLTKEEQAFQWQVTRDADAMAKVASAANQAVTSADQGLAVNNLPT-- 783
Qy 53 VFAAHKLPPA-----DSADGQAADVHNQAITLTIETASRLHFEGEPTATIA----- 100
Db 784 --PRAKYPEALGHVQQAADAKRQAIRDNALTAEPQADALRQVDAQAATAEAAINQNHNTN 841
Qy 101 DTFAKAEKLD--RLATTTSGALRATPFMASLLOQ---MQPAINKGDWLPAPLK--PLTPLI 155
Db 842 ATLAKADSDGVKATINDINPQPRSKPAANQALEQVAAAKRQAINNNQLTDEEKAQIQOV 901
Qy 156 SGALSGAMDQV-----GTKWMDRATGDLHYLSASPDRLHDMAASVKRHSPLSRQVL 208
Db 902 DQALANAKTVQAAANDNGVQNA--KTAGTTAINNINPQGTQKAQIAIAEASQKRL 960
Qy 209 DTGVAVQTSYARNVARTVLPALASRPVQV-----GAVDLGVSMAGGI----- 251
Db 961 QGRNDLTTEERNALADLTAKAQAADAVNQARNNTGVAGAKDNGVAGIQINPTAVVKP 1020
Qy 252 --AANAGFGRNLLSVQSRDHQRGALVLGLKKEPKAQLSEENDWL--EAYKAIKSAISYGA 309
Db 1021 DARN-----IDQAARDKEAFQANTKLTIDEEKAAAIKKYQDARDAKAIDRAGSNGD 1074
Qy 310 ALNA-----GKRWAGLPDMATDAMGAVRSLSVSSLTQNLGLA 349
Db 1075 VNNAVNQGAIAQAIKALDQSPSAKDTAKAIAQNAADAKKA--AITANNALTQEKA-- 1150
Qy 350 GGFAGVGLQEMATKNITDPATKAASV-----QLTNLAGSAAVFAGWTT 393
Db 1131 -----AAIKQVEDEAAKAQAADVADSRKADVDRAKQDQGLQKISDVPVAPQPKLN 1179
Qy 394 AALTTPDPAVKKAESFIQTVKST-----ASSTTGYVADQTVKLAKTVKDMGGEA--ITHGTAS 449

Db 1180 AIAAVDQAATDKKAVINNDTTLTQEEKEAIAIRKVDEEAKARQAINDATSNADVAAKQAO 1239
Qy 450 LRNTVNNLRQRP-----AREADIEEGGTA 473
Db 1240 GTQAINNVQPTPAKNAKAQAQVQAADA 1267
RESULT 3
S33441
EF protein - Streptococcus suis
C;Species: Streptococcus suis
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C;Accession: S33441
R;Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.
submitted to the EMBL Data Library, May 1993
A;Description: Repeats in an extracellular protein of wek-pathogenic strains are absent
A;Reference number: S33441
A;Accession: S33441
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1822 <SMI>
A;Cross-references: EMBL:X71880; NID:g298031; PIDN:CAA50714.1; PID:g298032
Query Match 6.1%; Score 147.5; DB 2; Length 1822;
Best Local Similarity 20.9%; Pred. No. 0.57;
Matches 106; Conservative 78; Mismatches 220; Indels 103; Gaps 19;
Qy 14 ATDSFRTASDASLASVRSVSSDQREINAIADYLTDHVFAAHKLPPADSADQQAADV 73
Db 1222 ARDAVELAKDKELAKEAIRTEEBEATKIVEKLAE---DTRKAIEDNPNLSDDEKQAEIKK 1278
Qy 74 HN---AQITLTIETASRLHFEGEPTATIAITFAKAEKLDRLATTTT-----GAL 120
Db 1279 LTDAVAKTLATIRDNADKRTQEAQAQALAD--LEKAKETQKIADKAAIDRLTILVKDDEL 1337
Qy 121 RATPPFAMASILLQYMPAINK-----GDWLPAPLKPLTLPILSGALSGAMDQ 165
Db 1338 EAT-----KQDAKNKIAKDAAAKAEIAIASNPNTLDAEKKFTTDAVDAEYAKANDA 1387
Qy 166 VGTWMDRATGDLHYLSASPDRLHDMAASVKRHSPLSRQVLDTGVAVQTSYARNVART 225
Db 1388 ISA-----ATTS-----PADVQKEEDAGVAAIAEDVLDAAKQAKKIAKAAAEKAEIGS 1437
Qy 226 VLPALASRPVQGVAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGAL---VLGLKDK 282
Db 1438 NPNTLDAEKKFTTDAVDAEYAKANDAIISAA---TSPADVQKEEDAGVAAIAEDVLDAAKQ 1494
Qy 283 EPKQALSEENDWLEAYKAIKSA-----SYSGAALNAGKRMAGLPDMATDAMGAVRSLSV 337
Db 1495 DAKNKIAKESD-----AAKSAIDANPNLTDAEKESAKKAVDADAKAATDAIDASTSPVE 1548
Qy 338 ASSLTQNLGLAGGAGVGVKGLQEMATKNITDPATKAASVOLTNLGSAAVFAGWTTAALT 397
Db 1549 AQSNEDK-----GVGSI-----AQDVLDAKQAKKIAKEVAAA-----KEAID 1588
Qy 398 TDPVAVKKAES--SFIQTVKSTASSTTGYVADQTVKL--AKTVKDMGGEAITHGTASLRNTV-- 454
Db 1589 ANPNLSDAEKEASKKAVDADAKATTDIDAIDASTSPVEAQSAAEDKG-----VGSIRQDVL 1641
Qy 455 -----NNLRQRPAREADIEEGGTAAPS 477
Db 1642 DAAKQDAKNKIAKESDAKSAIDANPN 1668
RESULT 4
T34852
Probable secreted protein - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T34852
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1999

A:Reference number: Z21559
 A:Accession: T34852
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 11156 <OLI>
 A:Cross-references: EMBL:AL035478; PIDN: CAB36606.1; GSPDB:GN00070; SCOEDB:SC2G5.19
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SC2G5.19

Query Match	6.1%;	Score 147;	DB 2;	Length 1156;
Best Local Similarity	23.1%;	Pred. No. 0.34;		
Matches 122;	Conservative	64;	Mismatches 231;	Indels 112;
				Gaps 21;

```

Oy 13 TATDSFRR-----SDASLASSVRSVSDOQREINLADLTGTHVPAHKLPEAD--- 63
Db 96 TAEDRRRVTLGMLSLSGMGMTASAAALAEAGE---AVAAFL-DGGRKTAELQDLRRNT 150
Oy 64 -SADGAAVDVHNAQITLIEFRASRLH-FEGETPATIADTFPAKEKLDRLATTTSGALR 121
Db 151 LSIONSARGRKMAQOTLSDGTSDALSVFLDTPFTANRDERVEVFTILAN----- 203
Oy 122 ATPFMAISLLOYMOPAINKGDWLPAPLKPFLPLISGALSG-AMOVGTXMMDBRATGDLH 179
Db 204 -----ASPEVAKYAGPALDEG-----TP---SAIHMFILIGQYIARARDETTATVD 246
Oy 180 YLSASPDR-----LHDMAASVYK-RHSPELARQVLDTG-----VAVQTSAR 220
Db 247 QLVAVEEGKRAQITTSDBRAVNASDKAEAAKAEALTAABAEAAAREVAKSAAAR 306
Oy 221 NAVRVLPLALASRPAYOGAVDLGVSMAGLAAANGPGRLLISVGRHQRGAL----- 275
Db 307 KRAAAKGAASAAKTAIVQAS-----SAHHNAKRBAFPAATTAQAQAATRGRAAALATYDA 361
Oy 276 VLGKLD-KEPRAQISEENDWLEAYKAKISASYSGAALNAGKRMAGLPIDMATDMGAVR 333
Db 362 VAAARDASKTKAARILAAEGARVYAAKARKRAQALAAQATATQAAAAAGISAAATARDSA 421
Oy 334 S-----LVSASLTLONGLALAGFAGVGKLOEMATKNITTPATKAA-VSGLTTLAASA 385
Db 422 AAQQAAVAAQASGAQAQSEBAVARAAAAAEADQAAR-----ATRAANPAOSLANTAYASA 474
Oy 386 AVFA-----GWTTLATTPPAVKTAESFIQDTYKSTASTTGYVAD 426
Db 475 AAAARKAADSAAHAEKADADAADAADAAGEADVYANKAKAWAADSV-AAAELLAAXVD 532
Oy 427 QTVKLAKTVKMGGAITHTGASLANTYNNLQRPARENDIEGGTAAS 475
Db 533 DARAFAAAREAEAEKLAH---DTEQSLAEAREMAAABEDAEARNA 578

```

```

RESULT 5
B90835
Probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain R1MD)
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C/Accession: B90835
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kuokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kunhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA. Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome
A/Reference number: A99629; MUID:21156231; PMID:11558796
A/Accession: B90835
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-971 <HA>
A/Cross-references: GB:BA000007, P1DN:BA935073.1, P1DN:G13361114, GSFDB:GN00154
A/Experimental source: strain O157:H7, substrain R1MD 0509952
C/Genetics:
C/Gene: ECe1650

```

```

Matches 107; Conservative 52; Mismatches 232; Indels 85; Gaps 15;

OY 19 RTADGASLASSVSVSVDQREINATADYITDHVPAAKLPAPDSADGQAAVDVHNQI 78
Db 115 RNASVVAQNTAAKSSADASTAREATNATD---AADSARAASTGQAASSAQSS 171
OY 79 TALETRASRLHPEGEFATADTFPAKAEKLDRLATTGSGLRATPPMAISILGYMPIAI 138
Db 172 SA-----GASTATKATASAAAESKSAATNAGAAKTSETNAAVSQ--QSA 219
OY 139 NKGDWLPAPLKLPLPLISGALSGAMDOVGTKMDRATGDLHYLSAPDRLDMAA--SV 196
Db 220 TSA-----STATTKASEAASSARDASASKEAKSS-ETSPASSASAASSATPAAGNSA 271
OY 197 K-----RHSPSLARQVLDTGVAQVQYSA--RNAVRTVLAPALASRPVQGVADLGV 245
Db 272 KAAKTSETNMASSSETAAEQSAAAGSKTAAALASASAISTSAGASASATPAAGSASAA 331
OY 246 SMAGGLAANAGFGRLLSVQSRDHQRGALVLGKDEPKQAQLSEENDWLEAVYAIKAS 305
Db 332 SSASTATTKAGEATEQASAAASSASAAKTSETNKASETSESSTKTAASASASAAASS 391
OY 306 YSGAA-----LNAQGRMAGLPLDMATDMAGVRSLSVASLTLQNGILALACGPAVGKL 358
Db 392 SASASKDEATQASAAKSSATTAATKATKAENGSTLAAQGSTTESAATRA----- 442
OY 359 QEMATKNTITDPA-----TKAAVSQLTNLGSAAVFAGWTTAALTDP-AVKAAS 407
Db 443 -ETAAKRAEDIASVALEDASTKKGIQVLSATNS-----TSESLAATPKAVKAAYE 494
OY 408 F-----IODIVK-----STASSTTGVAADQTVLAKTVKMGGEALTHITGAS 449
Db 495 LANGKTTAQDPTTKAQKGIQVLSATNSTSEMLAATPKSVKAAVDLANGKTTAQDAT 550

```

RESULT 6
C85693
probable membrane protein of prophage CP-933x Z1918 [imported] - Escherichia coli (strain
C1;Species: Escherichia coli
C1;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C1;Accession: C85693
R1;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A1;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A1;Reference number: A85480; MUID:21074935; PMID:11206551
A1;Accession: C85693
A1;Status: Preliminary
A1;Molecule type: DNA
A1;Residues: 1-973 <SNO>
A1;Cross-references: GB:AE005174; NID:G125114847; PIDN:AAGS6007.1; GSPDB:GN00145; UMGF:Z15
A1;Experimental source: strain O157:H7, substrain EDL933
A1;Genetics:
A1;Gene: Z1918

```

Query Match          6.0%; Score 145.5; DB 2; Length 973;
Best Local Similarity 22.5%; Pred. No. 0.33;
Matches 107; Conservative 52; Mismatches 232; Indels 85; Gaps 15;

QY      19 RTAGDASLASSVSVSSDQOREINAIADYITDHYPAAKLPPADSADGQAVDVHNAQI 78
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db       117 RNAAVAVANTTAARKSASDASTAREATHATD--AADSARAATSGAGQAASSAGSASS 173
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY       79 TALIETRASRLHPEGERPATIADTFAAAEKLDRLATTSGALRATPEPMASSLQYMPAI 138
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db       174 SA-----GTAFTATATEASKSAAAEESKSAATSGAATKSTETNAVSG--GSA 221
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      139 NKGDMLPAPLKEPLPLISGLSGAMDQVGTQMDRATGDLHYISASDRLLHDAQA--SV 196
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      222 TSA-----STATTKASEANASRPDASAGEAKKS--ETSSAASSASAASSATRAAGNSA 273
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      197 K-----RHSPSLARQVLDITGVAVQYTS--RNAVTVTLAPALASPPAVQAGVDLGV 245
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 274 KAAKTSNNAKSSATAEQSASARAAGSKTAAALASASASTSAGQASASATAAGKSAESA 333
Qy 246 SMAGGLAANAGFGRNLLSVQGRDRHORGGALVGLKDKPKAQLSENDWLRYAKIKAS 305
Db 334 SSASTATTKAGEATPQAGAAASASAAKTSNNAKASSETSAESSKTAASASSASAASAS 393
Qy 306 YSGAA-----LNAGKEMAGLPLDMATDAMGAVRSLVSSASLTQNLGALAGGAGVCKL 358
Db 394 SASAKDBEATQASAAKSATTTASTKATTAAGSATAAQAQSKSTAESATRA----- 444
Qy 359 QEMATKNITDPA-----TKAAVSQLTNLGSAAVFAGATTAAALTTP-AVKKAES 407
Db 445 -ETAAKRAEDIASAVALDEASTTKGIVQLSSATNS-----TSESAAATPKAVKAYE 496
Qy 408 F-----IQDTVK-----STASSTGYVADQTVKLAKTVKMGGBAITHTGAS 449
Db 497 LANGKYTAQDATTAQKIGIVQLSNATNSTSEMLAATPKSKVAAYDLANGKYTAQDAT 552
RESULT 7
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisel, C.; Gattung, S.
A:Submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: 221525
A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GEI>
A:Cross-references: EMBL:U80846; PIDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetic:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/
Query Match 6.0%; Score 144.5; DB 2; Length 2232;
Best Local Similarity 18.7%; Pred. No. 1.2;
Matches 99; Conservative 81; Mismatches 261; Indels 89; Gaps 11;
Qy 3 INRRVQPPVTATDSFRTPASDASLASSVRSSVSDQQRREINAIADYLDTHVFAAHKLP- 61
Db 1306 IGSTVTQPS-TVSGNSSGSTVTIGSSEASTSGSFKTSPSSISPVPT-----SSPIPST 1359
Qy 62 --ASADGQAAVDVHNAQITALI-----ETRASRLHFEGETP 96
Db 1360 TFASTSGSTISDVSSVSTTSLAPLSSLPSTVPSSTQSFSTSEGSKASSPVPSQTS 1419
Qy 97 ATIADTFAKELDRLATTTSGALRATPFAMASLLQYMQPAINK--GDWLPAPLKPLTPL 154
Db 1420 STPTNPTGSESSLTSLTISGTHQHTWKRAS--SGSTSPSTNSQTSQTSVTVMGSSSTSGV 1478
Qy 155 ISGALSGAMDQVGTQKMDRATGDLHYLSASPDRLHDMAA-----SYKRHP 201
Db 1479 STSSASSTQPMQSTSQGSSAGSTVASSTASPAASSTAPSSSTGTMSTSSSTSGVTSISES 1538
Qy 202 SLARQVLTGVAV-----QTSARNAVRTVLAPALASRPA 236
Db 1539 TTASASSQTSVTVMGSSSTSGVTSASSTQPMQSTSQGSSAGSTVASSTAGLVSTSTV 1598
Qy 237 VQGAVDLGVSMAGGLAANAGFGRNLLSVQGRDRHORGGALVGLKDKPKAQLSENDWLE 296
Db 1599 PSSCTMGSTSGTSGTISSESTTASASS---QTGSTVTMGSSSTSGVTSASSTQTPQ 1655
Qy 297 AYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSSASLTQNLGALAGGAGV 356
Db 1656 MSTSGGSSAGSTVASSTTGLVSTSTVPSSTGTMGSTSGTSGTISSESTA-ASASSQTG 1714
Qy 357 KLOEMATKNITDPAKAAVSQLTNLGSAAVFAGATTAAALTTPDPAVKKAESFIQDTVKST 416

Db 1715 STVTMGSSSTSGVTSASAGSQPMQSTSQSSAGSTVVSSTASPAASSTAPSTGTMST 1774
Qy 417 ASSTGTGYVADQTVKLAKTVKMG-----GBAITHTGASLRNTV 454
Db 1775 SSGTGTMGSSSTAASTTSHGTSTVTLGSSSTSSNQMSSTSQSSVGVSTV 1824
RESULT 8
C97523
hypothetical protein AGR_C.2490 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: C97523
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurolo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: C97523
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2155 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87140.1; PID:g15156408; GSPDB:GN00169
C:Genetic:
A:Gene: AGR_C.2490
A:Map position: circular chromosome
Query Match 5.7%; Score 138.5; DB 2; Length 2155;
Best Local Similarity 20.0%; Pred. No. 2.6;
Matches 119; Conservative 82; Mismatches 194; Indels 199; Gaps 23;
Qy 17 SFRTPASDASLASSVRSSVSDQQRREINAIADYL-----TD 51
Db 218 ALRLAEPETVASDRIMSVGOAVRREVSAMNDGIERTIARTATELTTLVHSEVNALERSVAD 277
Qy 52 HVFAAHKLPADSADGQAAVDVHNAQITALI-----ETRASRLHFEGETP 96
Db 278 NELRVSLVQELTAERDAIVN-HAERIRSSIVGAQEQIKEELSIIVGEELSMRIATTGEAF 336
Qy 97 ATIADTFAKA--EKLDRLATTTSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPL 154
Db 337 ASMDITSAALLEK-SRASTEAMGSL-----IAAKTENLQALNS----- 375
Qy 155 ISGA-----LSGAMDQVGTQKMDRATGDLHYLSASPDRLHDMAASVKRHS 202
Db 376 -SGSTISNEFDMRLHNLSTLDERGEVLLERFAIHASTLDSGVESLNSALEERQLNET 434
Qy 203 LARQVLTGVAVQTYASARNAVR-----TVLAPALASRPAVQAVD-- 242
Db 435 LSARSLEIN-----RNTERGQOVIGSLDVTLDKLTSTLEEKGLSFQSLQSTADDA 486
Qy 243 -LGVSMAAGL-----AANAGFGRNLLSVQGRDRHORGGALVGLKDKPKAQLSEE 291
Db 487 IMDLDRSLGLYEERMQATVQVNSAFDEHVAQPAFAFDQAGSL-----DSKLMESLARI 541
Qy 292 NDWL-----EAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSSASLTQNGLA 347
Db 542 NETVAGGSEALDTI-----LTSGLERIG-----STWTDQSLA 573
Qy 348 LAGGFAGVGKLOEM-----ATKNITDPAKAAVSQLTNLA-GSAAVF 388
Db 574 LA-----TALGTGQEMLENALESRTQAFSDAIGQRTAEITDAFTNSHAKITDTVLAERSNALF 630
Qy 389 AGWTTAALTTPPAVKKAESFTQDTVKST-----ASSTGYVADQTVKLAKTVKD 437
Db 631 GALSASQDRFDEALASRLAITGVSQGTAEHLAAMLDERAAAINSVVADVVERRITETLET 690
Qy 438 MGBAITHTGASLRNTVNN-LRORPA-----READIEGGTAASPSEI 479
Db 691 RAA-AITGAVSGIEDRISDTLSRSTAALHDVVGASERIAADTLDGRTAALSSAI 743

RESULT 9
AD2742
Conserved hypothetical protein Acu1348 [imported] - Agrobacterium tumefaciens (strain C5
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #ext_change 11-Jan-2002
C/Accession: AD2742
R/Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, I
erage, G.; Giller, W.; Grant, C.; Guenther, D.; Kutyavina, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; PMID:11743193
A/Accession: AD2742
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-2155 <KUR>
A/Cross-references: GB:AE008668; PIDN:AA142354.1; PID:g17739760; GSPDB:GN00186
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Acu1348
A/Map position: circular chromosome

Query Match 5.7%; Score 138.5; DB 2; Length 2155;
Best Local Similarity 20.0%; Pred. No. 2.6;
Matches 119; Conservative 82; Mismatches 194; Indels 199; Gaps 23;

QY 17 SPTASASLASSSVSSDQREINADYL-----TD 51
DB 218 ALRIAEPEYASDRIMVGAVREVSAMNDGERTIARATELETLVHSEVNALEERSYAD 277
QY 52 HVFAAHKLPADSGAAYDVHNAOITALI-----ETRSRLHFEGETP 96
DB 278 NEURVRLVQELTHERAIYA-HAERTRSSIVGAQEOIKESLTVSELSRLITTTGEAF 336
QY 97 ATADTPAKA-EKLDRLATTTGALRATPFPAMASLQYQMPAINKGDWLPAPLKPPLPL 154
DB 337 ASMDTSAALLEX-SASTEAMGSL-----IAKTEENLQALNS----- 375
QY 155 ISGA-----LSGANDVGTOMDRATGDLHYLSAPRLDAMAASVGRHSPS 202
DB 376 -SGSTISNEPDMRLHNLSTLDERGEVILRFIAHASTLDGVSLSALSEERTQLNET 434
QY 203 LARQVLDTVAVQVTSARNV-----TVLAPALASRPVQAGD-- 242
DB 435 LSRASLELN-----KNIERGQOVTIGSLDTVLKLTSTLEKGLSFROSLQSTADDA 486
QY 243 -LGVSNAGL-----AANAGFGNRLSVQRDHORGALVGLKDKPEKAOUSEE 291
DB 487 IMOLDLRSGLYBERMQATVGVNSAFDEHVAQFASAPDQRAQSL-----DSKLMESLARI 541
QY 292 NDMT-----EAYKAIKASYSYGALINAGKRNAGLPDMATDMGAVRSLVSASSITONGLA 347
DB 542 NEIVAGSGSEALDT-----LTSGLERIG-----STWTDQSLA 573
QY 348 LAGGFAVGLQEM-----ATKNITDPATKAASQTLNLA-GSAAVF 388
DB 574 LA---TLGIGQEMLENALESRTAFSDAIGORTAETIDAFTHSHAKIDVLAERSVALF 630
QY 389 AGWTTAALTTPDAVKAESFIQDTVKST-----ASSTGYVADQTVLAKTVKD 437
DB 631 GALSASQDRDEBALASRLAITSVSGTAHEHLAMLDERRAALINSVADVERRTLETLET 690
QY 438 MGGGAIHTTGASLNTVNN-LRQPPA-----READIEGGGTASPSSEI 479
DB 691 RAA-AITGAIVSGIEDRISDTLESTALAHVVGASERIRIADTDLGRTAALSAI 743

RESULT 10
T35746
Hypothetical protein SC7H2.15c - Streptomyces coelicolor
C/Species: Streptomyces coelicolor

C/Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #ext_change 05-Nov-1999
C/Accession: T35746
R/Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrrell, B.G.; Rajandream, M
submitted to the EMBL Data Library, August 1999
A/Reference number: Z21588
A/Accession: T35746
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-597 <SAU>
A/Cross-references: EMBL:AL109732; PIDN:CA852057.1; GSPDB:GN00070; SCOEDB:SC7H2.15c
A/Experimental source: strain A3(2)
C/Genetics:
A/Gene: SCOEDB:SC7H2.15c

Query Match 5.6%; Score 136; DB 2; Length 597;
Best Local Similarity 21.5%; Pred. No. 0.66;
Matches 116; Conservative 75; Mismatches 226; Indels 122; Gaps 21;

QY 2 HINRVQPEPVATDSFRTASDASLASSSVSVSSDQREINADL-YLTDHVFAAHKL 59
DB 4 HFDRLRLKAAVTTT-----VAAVVAALASQAPDVTADGNGRQTADNAPTSDT 53
QY 60 PPADSADGQAAV-----DVHNAQITALLIETRASRLHFEGETPTIADTFPAKAKLDRLAT 114
DB 54 PPEESATGNSRYTTDLPPLNSPSPAPTTGTPASRGASEAGIPATVDAVKKASELRRAK 113
QY 115 TTSGALRATPFPAMASLQYQMPAINKGDWLPAPLKPPLISAL-----SGA 162
DB 114 PGCN-----LFWQLAAIIGKVESGQARGVANDANGTTIGRIIPQDONGFALIKDTDNGV 169
QY 163 MDQVGTOMDRATGDLHYLSAS-----PRLDAMAAS--VGRHSPSLA 204
DB 170 YD--GMSVYDNAVGPQGFISTWAMAGRDGNSDKEDPNNVYDAAAGHYLCRNSMDLA 227
QY 205 RQVLDTVAVQVTSARNVRLTAPALASRPVQAGVLDGVSMAGLAAAGFGRNL-- 261
DB 228 DQA-DIKRALISTYNSQDYHTVLSWLEYRKQTHBIPODTGSLPGRSDATPGRSPG 286
QY 262 --LSVGRDHQRGALV-LGLKDKPEKAOUSEENDLBAVKAIKSASYSGALINAGKRNA 318
DB 287 RGTGASSRPOSPTATPSPGRSPSKPPA-----SCRPTDPSPT 326
QY 319 GLPLDMATDMGAVRSLVSASLTONGLAGFAGVGLQENATKNITDPATKAAS-- 376
DB 327 S-PSTPTDPAPEPTPTPTPDVHLENAGATGFSALAG-DFAERISTAEKTAGAVG 384
QY 377 ----QLTNLAGSAVPAW-TTAAITTD-----PAVKAESFIQDTVKSTASST-- 421
DB 385 KVRERFTVLGDDTATFTGGEKVAAVTTNASGVAVALVAGETGFTVATVIGRTVPG 444
QY 422 -GVVADQTVLAKTVDMGSEALITHGASLRNTVNNLRQRPAREADIEEGGTASPSSEI 479
DB 445 LDYKATVTERVA-----DALSTGET-----EPCTPGEGRADPLEV 481

RESULT 11
E95206
Cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain T
C/Species: Streptococcus pneumoniae
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #ext_change 03-Aug-2001
C/Accession: E95206
R/Fetellin, H.; Nelson, K.E.; Paulsen, I.T.; Bisen, J.A.; Read, T.D.; Peterson, S.; Hei
on, J.D.; Umeyam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holzapfel,
neon, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A/Reference number: A95000; MUID:21357209; PMID:11463916
A/Accession: E95206
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-4776 <KUR>
A/Cross-references: GB:AE005672; PIDN:AAK75846.1; PID:g14973269; GSPDB:GN00164; TIGR:SP4

QY 238 QGAVDLGVSMAGLAANAGFENRLLSVSRDHQCGALVLCQKPEKQAQISENDMLEA 297
 Db 1127 LEVTDIASLSMAQIA-----GAGVAGL--TTDQIALS-----ISQ 1163
 QY 298 YKAIKSASYSGAALNAGKRMAGPLD-----MATDAGAVRSVSA--SSLTQNGL 346
 Db 1164 VEAITSQI--MALS--KQIALSADLAIITTAEMMAISGALSGLPASTITSLITRQI 1220
 QY 347 ALAGGFAGV---KQEMATKNITDPATKAASOLT-----NLG---SAVPAGW 391
 Db 1221 AALGAAVSGLTQDQIALGTGV--DALTNQIGALTSKQYALSVSGISLSAQIAL 1279
 QY 392 TT--AALTDPAYKKAESFQ--DYKSTASSTTGYA---DQYKLAATVDMGSEA 442
 Db 1280 STAGVAGLTQDQIALSTQVQALTSVQIALSSKOIALGADLDITTAIASISGSA 1339
 QY 443 ITHGASLRNTVNNLRQRPAREADIEEGTAA 474
 Db 1340 VAGLSAS---TIASL-----TTAQIALAGTAA 1363

RESULT 14

F87386
 Conserved hypothetical protein CC1106 [imported] - Caulobacter crescentus
 C/Species: Caulobacter crescentus
 C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C/Accession: F87386
 R./Merman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.;
 n, J.; Embleave, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Frazer, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A./Title: Complete Genome Sequence of Caulobacter crescentus.
 A./Reference number: A87249; MUID:21173698; PMID:11259647
 A/Accession: F87386
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residue: 1-2479 <STO-
 A/Cross-references: GB:AE005673; NID:913422414; PIDN:AAK23090.1; GSPDB:GN00148
 C/Genetic:
 A/Gene: CC1106

Query Match 5.6%; Score 135.5; DB 2; Length 2479;
 Best Local Similarity 23.1%; Pred. No. 4.8;
 Matches 113; Conservative 67; Mismatches 211; Indels 99; Gaps 19;
 QY 13 TATSPRTASDASLASSVRSVSSDOOREINAI--ADYLTDFVFAHKLPPADSADGQAAV 71
 Db 1766 TQIALQTSQALALITQAGTSLTQGLALNAVQALITQAFSALDATQVQALTTQVK 1825
 QY 72 DVHNAQTALLETBASRLHFEGETPATIADTFAKAEKLDRLATTSQALRATPPAMASLL 131
 Db 1826 GLTNSQSALATTTDVQI-----ADT-----QVGLSATQGLTATASALDAT 1870
 QY 132 QYMPALINKDMLPA--PLKPLTPLISGALS-----GAMDVGTRMMDR----- 173
 Db 1871 QTAALTTQVKGATATQKGLTTDDIDGLADTVQALSAQGLTATGSPESLNAVAV 1930
 QY 174 -----ATGDLHTLASPDRLHDMAASVSRHSPSLARQVLDGVAVQT 216
 Db 1931 LSAQTQIGLTADQKSGSTTDIGELSTT--QITALTADQLKNLSATINISAAADQVAAIS 1988
 QY 217 YSARNAVTVLAPLAPASRAVQ--GAVDLGVSMAGLAANAGFENRLLSVSRDHQCGAL 275
 Db 1989 NTOIALSTTQKGLSTVDIDELSATQVGLSAAQIALASTQIGELSTTQ----- 2039
 QY 276 VLGKDKPEKQAQISENDMLEAYK--AIKASYSGAALNAGKRMAGPL---DMATDAM 329
 Db 2040 IGGISATQIGAVSATVNSLNAVQVGLTTQVKGTLV---AQLAGSTTNIGDLADTV 2096
 QY 330 GAVSLVASASTQNGALAGGPA-----GVYKQEMATKNITDPATKA--AVSOLTNLG 383
 Db 2097 GALLS-----TQGLATATGFTSTNGTQVAAINLSQKGLTTTQKGLTTTIDISELAD 2149

QY 384 S--AAVPAQTALTTDPAYKKAESFIQDTVKSTASSTT---GYVADQVTKLAKTYKM 438
 Db 2150 TOVAAITAAQIALSATVNSALNA-----TQVATLTQVKGTLVDQKGL--TTTDDI 2200
 QY 439 GGEALHTTGA 448
 Db 2201 GELANTQIGA 2210

RESULT 15

AB1122
 Hypothetical protein NME1108 [imported] - Neisseria meningitidis (strain MC58 serogroup
 C/Species: Neisseria meningitidis
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C/Accession: AB1122
 R./Tetzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scallato, V.; Maignan, V.; Pizze, M.
 Science 287, 1809-1815, 2000
 A./Authors: Grandi, G.; Sun, L.; Smith, H.O.; Frazer, C.M.; Moxon, E.R.; Rappapoli, R.; V.
 A./Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A./Reference number: AB1000; MUID:20175755; PMID:10710307
 A/Accession: AB1122
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residue: 1-718 <TET>
 A/Cross-references: GB:AE002460; GB:AE002098; NID:97226335; PIDN:AAF41499.1; PID:9722633
 A/Experimental source: serogroup B, strain MC58
 C/Genetic:
 A/Gene: NME1108

Query Match 5.6%; Score 135; DB 2; Length 718;
 Best Local Similarity 22.4%; Pred. No. 0.97;
 Matches 119; Conservative 76; Mismatches 224; Indels 112; Gaps 25;
 QY 14 ATDSFRTASDASLASSVRSVSSDOOREINAIADYLTDFVFAHKLPPADSADGQAAV 73
 Db 161 ATBGAQOIKD--LALVELKRGCTHDKALDLISGMVTTGTFQTKNEAQAFAALASE 218
 QY 74 HNAQITA-LIET-----RASRLHFEGETPATIADTFAKAEKLDRLATTSQALRATP 124
 Db 219 GSGEDTKAKLTKLKGDSKSDYQLGHEVLSGLDSTFVRDVRRELPSLSAAQAGM 278
 QY 125 PAMAS--LLQYMPALINKDMLPAPL-----KPLTPLISGAL-----SGAM 163
 Db 279 NGVGGDLVLSLQSANCKG--SPAELATVQNLSTKSLSPDTIGRLKKRANPRDPKGV 337
 QY 164 DQVGTKMMDRATD--LHYLSASPDRLHDMAASVSRHSPSLARQVLDGVAVQVTSAR 220
 Db 338 DWIGSVVQKQNGENNAVQVLS---RLADMVLVQKQYQYKRRAAAGDTAAEQAMLK 393
 QY 221 NAVRTVAPLAPASRAVQAVDL-----GYSMAGLAANAGFENRLLSVSRDH 269
 Db 394 GALLAQILPLQKQGLLAATDMTQIREVNASLAGVTLQNGKIAKNEA--RMLSAQAQOE 452
 QY 270 QRGALVTLGKDKPEKQAQISEE-----NDWLBAVKAIKSASYSGAALNAGKRMAGPLDM 324
 Db 453 Q-----QESLAMEBSLTGLTVDMETSFKCL--AAEYFNATL----- 487
 QY 325 ATDAMGAVSLVASASTTQNGALAGGPAVQKQEMATKNIT--TDPATKAASQLTNLG 383
 Db 488 ---ALQALTTAAATASAM--LITAGGKQAGFLKDVGSYALGMGRKASAGVAAATAG 542
 QY 384 SAAVPAQTAA---LTTDPAYKKAESFIQDTVKSTASSTTGYVADQVTKLAKTYKM 440
 Db 543 GKLL--SWGKSASGGLNNNVLVPRAGLGLMLLYSSS---LGQGLP---KGLRG 589
 QY 441 EALTHTGASLRNTV--NNLRQRPAREADIEEGT---AASPSEIIPRPKMS 486
 Db 590 ---TKTTPENINLKNNGIRFEPAPKREQARGVQVLAAPSAQPTDKMS 637

Fri Jan 3 18:00:36 2003

us-09-825-414-7.rpr

Page 8

Job time : 29.9825 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:19:06 ; Search time 8.50874 Seconds
(without alignments)
1084.696 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407
Sequence: 1 MHINSAQOPGVAMSEFRT.....EEGGSIAFSRSTPQLRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2407	100.0	487	10	US-09-825-414-66
2	1869.5	77.7	486	10	US-09-825-414-7
3	147	6.1	1463	9	US-09-971-536-69
4	146.5	6.1	2122	9	US-09-813-214A-9
5	140	5.8	2285	10	US-09-932-183A-2
6	138	5.7	2293	9	US-09-738-626-6239
7	132	5.5	1026	10	US-09-379-931-7
8	130.5	5.4	2368	10	US-09-815-242-5635
9	130.5	5.4	2368	10	US-09-815-242-15389
10	129.5	5.4	2076	10	US-09-815-242-5815
11	129.5	5.4	2186	10	US-09-815-242-15913
12	129	5.4	2344	10	US-09-815-242-15713
13	127.5	5.3	2478	10	US-09-815-242-5816
14	127.5	5.3	2478	10	US-09-815-242-12967
15	125.5	5.2	1222	10	US-09-137-531-15
16	125.5	5.2	1252	10	US-09-137-531-9
17	125	5.2	2086	10	US-09-815-242-5639
18	125	5.2	5795	10	US-09-815-242-12610
19	124	5.2	628	10	US-09-841-786-5

20	124	5.2	3241	10	US-09-841-786-1	Sequence 1, Appli
21	123.5	5.1	2434	10	US-09-815-242-5835	Sequence 5835, Ap
22	123.5	5.1	6281	10	US-09-815-242-12996	Sequence 12996, A
23	120.5	5.0	810	9	US-09-712-363-281	Sequence 281, App
24	119.5	5.0	778	10	US-09-815-242-10728	Sequence 10728, A
25	119.5	5.0	1029	10	US-09-815-242-5885	Sequence 5885, Ap
26	119.5	5.0	1048	10	US-09-815-242-13083	Sequence 13083, A
27	119	4.9	600	10	US-09-287-849-22	Sequence 22, Appl
28	118	4.9	596	9	US-10-063-547-100	Sequence 100, App
29	118	4.9	596	12	US-10-066-867-100	Sequence 100, App
30	118	4.9	596	12	US-10-052-586-310	Sequence 310, App
31	118	4.9	647	10	US-09-841-132-436	Sequence 436, App
32	117.5	4.9	1215	10	US-09-815-242-5908	Sequence 5908, Ap
33	117.5	4.9	1269	10	US-09-815-242-13113	Sequence 13113, A
34	117.5	4.9	3158	10	US-09-815-242-12611	Sequence 12611, A
35	114	4.7	573	10	US-09-828-574-13	Sequence 13, Appl
36	113	4.7	729	10	US-09-287-849-2	Sequence 2, Appli
37	113	4.7	1054	10	US-09-798-042-87	Sequence 87, Appl
38	112	4.7	660	12	US-10-007-693-139	Sequence 139, App
39	112	4.7	1879	9	US-09-971-536-70	Sequence 70, Appl
40	111	4.6	1488	9	US-09-738-626-5495	Sequence 5495, Ap
41	110	4.6	792	9	US-10-025-380-1127	Sequence 1127, Ap
42	109.5	4.5	4999	9	US-09-976-059-14	Sequence 14, Appl
43	108.5	4.5	764	10	US-09-815-242-5143	Sequence 5143, Ap
44	108	4.5	458	9	US-09-738-626-6104	Sequence 6104, Ap
45	108	4.5	596	10	US-09-287-849-26	Sequence 26, Appl

ALIGNMENTS

```
RESULT 1
US-09-825-414-66
Sequence 66, Application US/09825414
Patent No. US20020083489A1
GENERAL INFORMATION:
APPLICANT: Collier, Alan
APPLICANT: Alfano, James R.
APPLICANT: Charkowski, Amy O.
TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
FILE REFERENCE: 19603/3243
CURRENT FILING DATE: 2001-04-03
PRIORITY FILING DATE: 2000-04-03
PRIORITY FILING DATE: 2000-04-03
PRIORITY FILING DATE: 2000-04-03
PRIORITY FILING DATE: 2000-08-11
PRIORITY FILING DATE: 2000-08-11
PRIORITY FILING DATE: 2000-11-17
NUMBER OF SEQ ID NOS: 91
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 66
LENGTH: 487
TYPE: PRT
ORGANISM: Pseudomonas syringae pv. tomato
US-09-825-414-66
Query Match 100.0%, Score 2407, DB 10, Length 487,
Best Local Similarity 100.0%, Pred. No. 2.9e-180;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 181 LSTSPDKLHDAMASVKRHSPPALGRQVDMGIATVTFSSALNVVTRTVLAPALASRPSVOGA 240
 DB 181 LSTSPDKLHDAMASVKRHSPPALGRQVDMGIATVTFSSALNVVTRTVLAPALASRPSVOGA 240
 QY 241 VDFGVSTAGGLVANAGFGRMLSVQSRDQLRGGAFLVGMKDKPEKKAALSEETDMLDAYKA 300
 DB 241 VDFGVSTAGGLVANAGFGRMLSVQSRDQLRGGAFLVGMKDKPEKKAALSEETDMLDAYKA 300
 QY 301 IKASYSYGAALNAGKRMAGLPLDVATDGLKAVRSLSVSATSLTKNGLALAGGAGVSKLOK 360
 DB 301 IKASYSYGAALNAGKRMAGLPLDVATDGLKAVRSLSVSATSLTKNGLALAGGAGVSKLOK 360
 QY 361 MATKNIITDSATKAAVSQSLNVLGSGVGFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420
 DB 361 MATKNIITDSATKAAVSQSLNVLGSGVGFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420
 QY 421 TSYVADQTVKLAKTVKDMGSEAISSGTASLRSTVNNLRHRSAPADIEEGGISAFSRSET 480
 DB 421 TSYVADQTVKLAKTVKDMGSEAISSGTASLRSTVNNLRHRSAPADIEEGGISAFSRSET 480
 QY 481 PFQLRRL 487
 DB 481 PFQLRRL 487
 RESULT 2
 US-09-825-414-7
 ; Sequence 7, Application US/09825414
 ; Patent No. US20020083489A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Collmer, Alan
 ; APPLICANT: Alfano, James R.
 ; APPLICANT: Charkowski, Amy O.
 ; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
 ; TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES
 ; FILE REFERENCE: 19603/3243
 ; CURRENT APPLICATION NUMBER: US/09/825,414
 ; CURRENT FILING DATE: 2001-04-03
 ; PRIOR APPLICATION NUMBER: 60/194,160
 ; PRIOR FILING DATE: 2000-04-03
 ; PRIOR APPLICATION NUMBER: 60/224,604
 ; PRIOR FILING DATE: 2000-08-11
 ; PRIOR APPLICATION NUMBER: 60/249,548
 ; PRIOR FILING DATE: 2000-11-17
 ; NUMBER OF SEQ ID NOS: 91
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 7
 ; LENGTH: 486
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas syringae
 US-09-825-414-7

Query Match 77.7%; Score 1869.5; DB 10; Length 486;
 Best Local Similarity 78.4%; Pred. No. 2.4e-138;
 Matches 381; Conservative 33; Mismatches 71; Indels 1; Gaps 1;
 QY 1 MHINQSAQQPPGVAMESFRFTASDASLASSSVRSVSTTSCRDQLAITDYLKHHVFAAHRFS 60
 DB 1 MHINRVQPPVTATDSFTASDASLASSSVRSVSDQDQREINATADYLTDRHVFAAHLK 60
 QY 61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIETFAKAEKFDRLATTASSAF 120
 DB 61 PADSAGGAADVHNAQITATETRASLHFEGEPTATIDTFAKAEKLDRLATTISGAL 120
 QY 121 ENTPEAASVLYQMOPAINKGDWLTPLKPLTPLISGALSGAMDOGVTKMDRAGDLHY 180
 DB 121 RATPFAMASLLQYQMOPAINKGDWLTPLKPLTPLISGALSGAMDOGVTKMDRATGDLHY 180
 QY 181 LSTSPDKLHDAMASVKRHSPPALGRQVDMGIATVTFSSALNVVTRTVLAPALASRPSVOGA 240
 DB 181 LSASPDRLHDAMASVKRHSPPALGRQVDMGIATVTFSSALNVVTRTVLAPALASRPAVQGA 240

QY 241 VDFGVSTAGGLVANAGFGRMLSVQSRDQLRGGAFLVGMKDKPEKKAALSEETDMLDAYKA 300
 DB 241 VDLGVSWAGGLANAGFGRMLSVQSRDQLRGGAFLVGLKDKPEKKAALSEENDWLEAYKA 300
 QY 301 IKASYSYGAALNAGKRMAGLPLDVATDGLKAVRSLSVSATSLTKNGLALAGGAGVSKLOK 360
 DB 301 IKASYSYGAALNAGKRMAGLPLDVATDGLKAVRSLSVSATSLTKNGLALAGGAGVSKLOK 360
 QY 361 MATKNIITDSATKAAVSQSLNVLGSGVGFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420
 DB 361 MATKNIITDPATKAAVSQSLTNLAGSAAVFAAGWTTAALTDPVAVKKAESFIQDTVKSTASST 420
 QY 421 TSYVADQTVKLAKTVKDMGSEAISSGTASLRSTVNNLRHRSAPADIEEGGISAFSRSET 480
 DB 421 TGYVADQTVKLAKTVKDMGGEAITHGTASLRSTVNNLRHRSAPADIEEGGTAA-SPSEI 479
 QY 481 PFQLRRL 486
 DB 480 PFRPWR 485
 RESULT 3
 US-09-971-536-69
 ; Sequence 69, Application US/09971536
 ; Patent No. US20020159976A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Glenn, Matthew
 ; APPLICANT: Havukkala, Ilkka
 ; APPLICANT: Bloksberg, Leonard
 ; APPLICANT: Lubbers, Mark
 ; APPLICANT: Dekker, James
 ; APPLICANT: Christenson, Anna
 ; APPLICANT: Holland, Ross
 ; APPLICANT: O'Toole, Paul
 ; APPLICANT: Reid, Julian
 ; APPLICANT: Coolbear, Timothy
 ; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
 ; TITLE OF INVENTION: Using Them
 ; FILE REFERENCE: 1043c2
 ; CURRENT APPLICATION NUMBER: US/09/971,536
 ; CURRENT FILING DATE: 2001-10-02
 ; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
 ; PRIOR FILING DATE: 2000-08-08
 ; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
 ; PRIOR FILING DATE: 2000-11-28
 ; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
 ; PRIOR FILING DATE: 2001-08-08
 ; NUMBER OF SEQ ID NOS: 83
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 69
 ; LENGTH: 1463
 ; TYPE: PRT
 ; ORGANISM: Lactobacillus rhamnosus
 US-09-971-536-69

Query Match 6.1%; Score 147; DB 9; Length 1463;
 Best Local Similarity 20.8%; Pred. No. 0.0025;
 Matches 104; Conservative 64; Mismatches 241; Indels 92; Gaps 12;
 QY 1 MHINQSAQQPPGVAMESFRFTASDASLASSSVRSVSTTSCRDQLAITDYLKHHVFAAHRFS 60
 DB 714 INANSAADVTSQAKATSAASDAASVYSAQSIAGSHADNME-----IKSLASDAEKQS 768
 QY 61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATI-----AETFAKAEKFDRLATT 115
 DB 769 QIALAASKSAASASAAIAIV--ASSAASEASSAAAVSNADASANSAAAYDSYASE 826
 QY 116 ASSA-----FENTPFAASVLYQMOPAINKGDWLTPLKPLTPLISGALSGAMDOVG 167
 DB 827 ASAAASANDSGYATASFAASASAAAM-----SAAALSTAQVAAK 865
 QY 168 TKMDRARGDLHYLSTSPDKLHDAMASVKRHSPPALGRQVDMGIATVTFSSALNVVTR--T 225

Db 866 VAVSDAAAG-----SAAVAVASAGSDSKKQATTAATANSQALDINKISLT 913
Qy 226 VLAPALASRPVSG-----AVDFGVSTAGGLVANAGFGDRLSYQSRDQLRGA 274
Db 914 DVASGASSSSSEAGQASTATSAVASAASSASERAGSYAHQAGSSASAVQSSGAQHAS 973
Qy 275 FVIGMKDKPKAALSEETDMLDYKAIKASYSYGALINAGKRMAGLPDVATGGLKAVRS 334
Db 974 TAASAAASYPKDGIOSLASQAASEAKASNSAATSA-----AAVGFSAASDASEQAKT 1029
Qy 335 LVGATSLTKGKLAVAGVAVSLQKATKNITDSATKAVSQLSNLVGSGVFPAGWTTA 394
Db 1030 AASADVVASSAASASTANSASAAA-----SATKAGDSKAA-----AGFSSA 1069
Qy 395 GLATDPAVKAESEFIODPKVSTAS--STTSYVADQTVKLAKTVKMSGEAISTGASLRS 452
Db 1070 ASAASASAKAEAVASAAASASDDSVASASAAAGFPKASAAAGASASAAASASS 1129
Qy 453 TVNNLRHRSAPEDIEBGIS 473
Db 1130 AAAQTRGASSSASEGQAS 1150

RESULT 4
US-09-813-214A-9
; Sequence 9, Application US/09813214A
; Patent No. US20020177200A1
; GENERAL INFORMATION:
; APPLICANT: Tucker, Kenneth
; APPLICANT: Flohila, Laura
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE PROTEIN-106 POLYPEPTIDE, GEN
; FILE REFERENCE: 7969-089-999
; CURRENT APPLICATION NUMBER: US/09/813,214A
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: 08/968,685
; PRIOR FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 9
; LENGTH: 2122
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-813-214A-9

Query Match 6.1%; Score 146.5; DB 9; Length 2122;
Best Local Similarity 22.2%; Pred. No. 0.0045;
Matches 111; Conservative 69; Mismatches 181; Indels 139; Gaps 23;

Qy 32 RSVSTTSCRDLQAITDYLKHFVFAHRFSVIGSPERDALAHNEQIDALVETR---ANR 88
Db 377 KTLITIGAGTSLTD-----HNIGVQNDGKQVLAETLTSLKRVTTENTLTANE 427
Qy 89 LYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAASVLQYMPAINKGDMLATPL 148
Db 428 KVTVGKT-----RL-TTDKIGFTND---NMGIDSKPYLDKDTGIIHAGG 467
Qy 149 KPLTPLISGL-----SGAMQVGTKMDRARGDLHYLSTSPDKLHDA 191
Db 468 QKTITKLAVGDDAATYGLKKNQTAESALQTFVTKKVDKGNQD-----ANDS 517
Qy 192 MAVSVKHSFALRGQVDM-----GIAVOTFSAINVTFLVAPALASRPVQAVDFGV 245
Db 518 KITTVGNKPPDGTQVNTLTKLKGNGVDYTT-----ETNGVTYFGL 558
Qy 246 STAGGL-VANAGFGDRLSVQ---SRDQLRGAFVLGMDK--KEPPAALSEET---DW 294
Db 559 NONNGILGVNSTINNDGLSVGNTNSKQIOVADGITFTDISNSKPAIGIENTTRITRDG 618
Qy 295 LDVYKAIKASYSYGALINAGKRMAGLPDVATD---GLKAVRSLVGTSLSLTKGLALAGG 351
Db 619 I-----GFANNTGSLDANKRPLTPTGINAGGKELTVNQSALINPATNG----- 660

Qy 352 YAGVSLQKQKATKNITDSATKAAVSQLSNLVGSGVFPAGWTTAGLATDPAV-KKASFIQ 410
Db 661 -GOLDPMNRLSTANTEKSSAAITIKDLYNLQVPLTFAG-----DTGPNVTYKLEILK 713
Qy 411 DK-VKSTASST-----SYVADQ-----TVKIAKTVKMSGEAISTGASLRSYNNLRHR 460
Db 714 VKGKKTADDLTNNGINGVADSTDNSLTVKIAKTLSDLAVNTKTFLPASDKVTVDGNN- 772
Qy 461 SAPEADIEBGISAFGRSET 480
Db 773 ---TAKLQNGDGL-FSKQNT 788

RESULT 5
US-09-932-183A-2
; Sequence 2, Application US/09932183A
; Patent No. US20020127641A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases from Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FaalSeq for windows version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-932-183A-2

Query Match 5.8%; Score 140; DB 10; Length 2285;
Best Local Similarity 20.5%; Pred. No. 0.016;
Matches 128; Conservative 80; Mismatches 224; Indels 192; Gaps 26;

Qy 15 MESFRASDASLASSSVRSVSTTSCRDLQAITDYLKHFVFAHRFSV-IGSPERDALA 73
Db 251 IELYQROAQVNVQNLNTRYGSSWSSNRQAVQDYLN---AVNSLNVSTGSSNNIRSQIG 306
Qy 74 HNEQIDAL---VETRANRLYSEGETPATIAETP-----AKA 106
Db 307 LNNQFRELMSNAQTAAHQASSFG--AELTQFKSMSTYLISSGLFYGAISGLKENVSOA 363
Qy 107 EKFDRLATTAASSAFENTPFAASVLQYMPAINKGDMLATPLKPLTPLISGALSGAMDQV 166
Db 364 IEIDTLMNTIRRMANPEPKYNELL---QESIDLGLTSLKIKIDIIQMGDPGRMGFDES 420
Qy 167 GTXMDRARGDLHYLS-TSPDKLHDMAAVSVKHSFALGRQV-----VDMGIAVQTF 217
Db 421 ELSTLTKTAQVQLVNSVDLPDDTNTLTAAMLNFNIAANDSISIADKLNEVDNNVAVTTL 480
Qy 218 SALNVRTVLAAPALASRPVQAVDFGVSTA-----GLVAVN-----AFSGDRMLS 263
Db 481 DLANSIRK--AGSTASTFGEVLENDLGYTTAIASTRESGINIGNSLKITIFAIRIGNQSS 538
Qy 264 VOSRDQLRGAFVLGMDKPKAALSE---ETDML-DAYKAIKASYSYG----- 308
Db 539 IKALBOIGISVKTAGGEASASADLISEVAGKWDTLSDAQKQNTSIGVAGIYQLSRRNAM 598
Qy 309 ---AALNKG----- 315
Db 599 NNSFIQNNAKTAANGTSASWSEQKYADSLQARVNKLQNNFTFPAIAASDAFISDGLIE 658
Qy 316 --RMAGLPDVATDGLKAV---RSIVSATS-----LTRGKLALAGG-YAGVSL---QKXA 362
Db 659 FTQAGSLINASTGIVKSGVFLPPLLAAVSTATLLSKNTRTLASSLIIGTRAMQGETTLA 718

Qy 363 TKNITDSATKAASQ-----LSNIYGVSVFAGWTAGLATDPA-VKKA----- 405
 Db 719 TAGLEAGMTRAASVRVLKLTALRGLLVSTLVGGAFALGNALESLSIFABAKKADDFE 778
 Qy 406 -----BSFIQ-----DKVKSTASSTSYVADQ---TVKLAKT-----VK 436
 Db 779 OSQQTNVFAITNKDSTDKLQQYKELQVKESRLTSDSEQEYLVQVTOQLAQTFPALVK 838
 Qy 437 --DMSGEAISSTGASLRSTVNLR 458
 Db 839 GVDGQGNAILKTNKELEKAIENTK 862

RESULT 6
 US-09-738-626-6239
 ; Sequence 6239, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAOKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OKAZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: Patent in ver. 3.0
 ; SEQ ID NO 6239
 ; LENGTH: 2993
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-6239

Query Match 5.7%; Score 138; DB 9; Length 2993;
 Best Local Similarity 21.7%; Pred. No. 0.034;
 Matches 119; Conservative 85; Mismatches 219; Indels 126; Gaps 23;

Qy 7 AQQPGVAMESPTASDASLASSSVRSVSTTSCRDLOAITDYKHHVFAAHRFVSIGSPD 66
 Db 1025 AHTAPDVLGVRAWPAFAAVKSAVPGTDSASVVEGMLSLVHLEHI-----VLKSDV 1077
 Qy 67 ERDAALAHNEQIDALVETRAFLY-----SEGETPATTAETFA-KAEKFDRLATTAS 117
 Db 1078 PTDGALKVSATADEYVDYDGLGLVIRAEIADAEGLATLAERFAIRGRGNVARTNT 1137
 Qy 118 SAFE---NTPFAAASVLQYMQP-----AINKGDMLATPLKPLTLP-----ISGALSGAMD 164
 Db 1138 SALPTTVDTPRSARAVATVVAPESMRPAVTSGD-----RNFHVSDVNAASLAGLPG 1189
 Qy 165 QVGTQMDRARGDLHYLSTSPDKLHDAMAVSVKRH-----SPALGRQVDMGI-----AVQT 216
 Db 1190 VIVHGWTSAIGELTAGAFAFNDEQIQTAPAAKVVEVTATMLAPVLPGEBIEFVSERSAVDN 1249
 Qy 217 FSALNVRTVLAPALASRPSVOCADVFGVSTAGGLVA--NAGFGDRMLSVQRDQLRGA 274
 Db 1250 RFGMGEVTRVTA-----TVNGNL---VLTATAVVAAPSTFYAPFGGQISQG----- 1293
 Qy 275 FVLGKMDKEPKALSEETDLDLDAKAKSASYSYGAALNAGKRMAGLPLDVAATDGLKAVR- 333
 Db 1294 --MGMEARNRNSQARAIDWDRADAHTRNK-----LGFSIVEIVENNPVETVAGEKFFHP 1345

Qy 334 -SLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSATKAASQCLSNLVGSVGVFAGWT 392
 Db 1346 DGVLYLTQFTQVGNATL-GVAQIAEMREAHALNQRAYFAGHSVGEYNALAAAYAGVLSLES 1404
 Qy 393 T-----AGLATDPVAKKBSFIQ-----DKVKSTASSTSYVADQTVKLAKTVKD 437
 Db 1405 VLEIVYRGLTMRHLVDRDENGLSNYAALARPKNKMLGTADNVFDYVA-----SVSE 1456
 Qy 438 MSGE-----AISSTGASLRSTVNLRHRSAPEDIEE--GGISAF---SRS 478
 Db 1457 ASGEFLEIVNVNLAGLQYAVAGTQAGL-----AALRADVENRAPGGRAFILIPGI 1506
 Qy 479 ETPFQLRRL 487
 Db 1507 DVPPHSSKL 1515

RESULT 7
 US-09-379-931-7
 ; Sequence 7, Application US/09379931
 ; Patent No. US20020009792A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Smit, John
 ; APPLICANT: Bingle, Wade H.
 ; APPLICANT: NO. US20020009792A1lellini, John F.
 ; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULO
 ; FILE REFERENCE: 08106/002003
 ; CURRENT APPLICATION NUMBER: US/09/379,931
 ; CURRENT FILING DATE: 1999-08-24
 ; PRIOR APPLICATION NUMBER: US 08/614,377
 ; PRIOR FILING DATE: 1998-03-12
 ; PRIOR APPLICATION NUMBER: US 08/194,290
 ; PRIOR FILING DATE: 1994-02-09
 ; PRIOR APPLICATION NUMBER: US 07/895,367
 ; PRIOR FILING DATE: 1992-06-09
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 1026
 ; TYPE: PRT
 ; ORGANISM: Caulobacter crescentus
 US-09-379-931-7

Query Match 5.5%; Score 132; DB 10; Length 1026;
 Best Local Similarity 21.0%; Pred. No. 0.022;
 Matches 101; Conservative 75; Mismatches 166; Indels 140; Gaps 23;

Qy 22 SDASLASSSVRSVSTTSCRDLOAITDYKHHVFAAHRFVSIGSPDERDAALAHNEQIDAL 81
 Db 42 SDAALNTLKLNVSTTAVAIQT-----YOFFTVGAPSAAG-----LDLFL 81
 Qy 82 VETRA-----RLYSEGETPATTAETFAKAEK-----DRLATTASSAFENTPFAA--- 128
 Db 82 VDSITNTDNLDAVYK-----FAQENRFINFINLATGAGAG--ATAFAAYTG 129
 Qy 129 -SVLYQYQPAINK--GWLATPLKPLPLISGALSGAMDQVGTQMDRARGDLHYLSTSP 185
 Db 130 VSYAQTVATAYDKIIGNAVAT-----AAGVDVAAAVAFLSRQANIDYLT--- 173
 Qy 186 DKLHDAMAVSVKRHSVPSALGRQVDMGIATVQTFSSALNVRTVL---APALASRPSVQGVND 242
 Db 174 -----AFVRANTPTTAAADIDLAVKAAIGITLNAATVSGIGYATATAAAMINDLSD 225
 Qy 243 PGVST--AGGL-----VANAGFGDRMLSVOS--RDQLRG-----GAFVLGMKD 281
 Db 226 GALSTDNAGVNLFTAYPSSGVSGSTLSLTGTDTLTGTANNDTFVAGEVAGAAATLVGD 285
 Qy 282 KEPKAAALSEETDLDLDAKAKSASYSYGAALNAGKRMAGLPLDVAATDGLKAVRSLVSATSL 341
 Db 286 TLSGGAGTDVLNWNVOA-----AAVTA-----LPTGVITISGIETM-NVTSGAAI 327
 Qy 342 TKNGLALAGGYAGVSKL-----QKMATKNITDSATKAASQCLSNL-VGS 384


```

Db      328  TLN---TSSGVTGLTALNTNTSGAAQTVTAGAGQNLTAATTAAGAAANVAADGRANVTAS 384
Qy      385  VGVFAGWTGLATDPVAKAESFIQDKVASTSTSYVADQVVKLAKTVKMGSGEALIS 444
Db      385  TGTSTGTTTG-----ANSAASGVSVSVANSSTTTTGAILA-VTGGTAIVTAQTAGNAVN 438
Qy      445  ST 446
Db      439  TT 440

RESULT 8
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5635
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5635

Query Match      5.4%; Score 130.5; DB 10; Length 2368;
Best Local Similarity 20.2%; Pred. No. 0.093;
Matches 99; Conservative 64; Mismatches 223; Indels 105; Gaps 17;

Qy      4  NOSAQQPGVAMESFRTAS-----DASLASSSVRSVSTTSCDDLAITDYLKHHVFAA 56
Db      702  NASLODEKDVANDKIGIKETKAIKDIDAATTNAQVEAIKTKAINDINQTAATTAKAAL 761
Qy      57  HRSFVSGSPDERAAL---AHNEQIDALVETRANRLYSEGETPATIAETPAKAEKFDRLA 113
Db      762  EEFDEVVQAQIDQAPLNPDTTNEVEVAIE---RINAAKVSGVKAIEATTTAQDLERVK 817
Qy      114  TTASSAENTPFAAASVLQYMQPAINKGDMLATPLKPLTPLISGALSGAMDQVGTCKMDR 173
Db      818  NEBISKIEN---ITDSTQTKMDYNEVKQAATAKTQNTVSNATN---EEV----- 863
Qy      174  ARGDLHLSTSPDKLHADAMVSVGRHSPPALGROYVDKGINVQTSALNVRTVLAPALAS 233
Db      864  AENDAAVEAAQKQGLHIQYVKSQGEVADTKSKVLDKINAIQT----- 906
Qy      234  RPSVQAGVDFGVSTAGGLVANAGFGDMLSVQSRDQLRGCAFVLGMDKQEPKALSEETD 293
```

```

Db      907  QAKVKAADTEVENA-----YNTRKQEIQNSN-----ASTTEKQAAYTE--- 946
Qy      294  WLDPAVAIKSASVGAALNAGKRMAGLPLDVAT---DGKAVSLVSATSILTNGGALAG 350
Db      947  LDTKQGEARTNIDDAANTS-----DVTAKNGINAIQVQAATTKSDADAEIA 996
Qy      351  GYAGVSKL-----QKATKNITDSATCAAVSOLNLSVGVFPAGWTT-----A 394
Db      997  QKASERKTAIEAMNDSTTEEQAAKQKVDQAVVTANADIDNAAANDVDNAKTTEATIA 1056
Qy      395  GLATDPAVK-KAESFIQDKYKST-----ASSTSYVADQTVKLAKTVKMGSGEALIS 444
Db      1057  AITPDANVPEAKQAIDRVQAOETAIDANNGATTEKKAAKQOVTEKTTADTAIDG-A 1115
Qy      445  STGASLRSTVN 455
Db      1116  HTNAEVEBAKN 1126

RESULT 9
US-09-815-242-12389
; Sequence 12389, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12389
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12389

Query Match      5.4%; Score 130.5; DB 10; Length 2368;
Best Local Similarity 20.2%; Pred. No. 0.093;
Matches 99; Conservative 64; Mismatches 223; Indels 105; Gaps 17;

Qy      4  NOSAQQPGVAMESFRTAS-----DASLASSSVRSVSTTSCDDLAITDYLKHHVFAA 56
Db      702  NASLODEKDVANDKIGIKETKAIKDIDAATTNAQVEAIKTKAINDINQTAATTAKAAL 761
Qy      57  HRSFVSGSPDERAAL---AHNEQIDALVETRANRLYSEGETPATIAETPAKAEKFDRLA 113
Db      762  EEFDEVVQAQIDQAPLNPDTTNEVEVAIE---RINAAKVSGVKAIEATTTAQDLERVK 817
Qy      114  TTASSAENTPFAAASVLQYMQPAINKGDMLATPLKPLTPLISGALSGAMDQVGTCKMDR 173
```

Db 818 NEEISKIEN-----ITDSTQTKMDAYNEVKQATARTKQATVSNATN-----EV----- 863
Qy 174 ARGDLHYLSTDPKLDHDAVSVKHSHPALGQVVDGMIAVQTFSSALNVVTVLAPALAS 233
Db 864 AEAADAEEAAQKQGLHDIQVVKSKQEVADTKSKVLDKINAQT----- 906
Qy 234 RPSVQAVDFGVSTAGGLVANAGFGRMLSVQSRDLRGGAFLVGMKDKPKAALSEETD 293
Db 907 QAKVPAADTEVENA-----YNTRKQEIQNSN-----ASTTEKQAAYTE--- 946
Qy 294 WLDAVKAIKSASYSAAALNAGKRMAGLPLDVAT-----DGLKAVRSLSVATSLSLTGNGLALAG 350
Db 947 -LDTKKQEAFTNLDANTNS-----DVTAKDNGIAALINQVQAATYKSKDAKAEIA 996
Qy 351 GYAGVSKL-----OKMATKNIITDSATKAASVQSLNLSVGVGFAGWTT-----A 394
Db 997 QKASERKTAIEAMNDSTTEEQQAADKVDQAVVTANADIDNAAANTDVDNAKTTNEATIA 1056
Qy 395 GLATDPVK-KAESFIQDKVST-----ASSTTSYVADOTVKLAKTVKMSGEAIS 444
Db 1057 AITPDANKPTAKQAIADKQVQAEATAIDANNAGATTEKAAKQVQTEKTTADTAIDG-A 1115
Qy 445 STGASLRSTVN 455
Db 1116 HTNAEVEAAKN 1126

RESULT 10

US-09-815-242-5815
; Sequence 5815, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5815
; . LENGTH: 2076
; TYPE: PRN
; ORGANISM: Staphylococcus aureus
US-09-815-242-5815

Query Match 5.4%; Score 129.5; DB 10; Length 2076;
Best Local Similarity 19.3%; Pred. NO. 0.093;
Matches 106; Conservative 91; Mismatches 201; Indels 151; Gaps 22;
Qy 20 TASDASL-----ASSSVRSVSTTSCRDQIAITDYLKHVFAAHFSPVIGSPDPERDAALAHN 75

Db 930 TAANTNILNANTNADVEQVKTNATQIGQIATPATK-----VKTDAKNA----- 972
Qy 76 EQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQ 135
Db 973 --IDKSAETQHTTIFNNND--ATLEEQAAQQLLDQAVATAKQIN-----AADTNQEVA 1023
Qy 136 PAINKGDWLATPLKPLTPLISGALSAMDQVGTK-----MMDRARGDLHYLSTSPDKLHDA 191
Db 1024 QAKDQGTQNIWVQPAT-----QVKTDTRNVVNDKAREAITNINATTGATREE 1071
Qy 192 MAVSVKHSHPALGQVVDGMIAVQTFSSALNVVTV-----VLAPALASRPSVOGA----- 240
Db 1072 KQEAIRNVVTLKRALTDIGV-TSTTANVNSIRDDAVNQIGAVOPHVTKKQTATGVNLNDL 1130
Qy 241 -----VDFGVSTAGGLVANAGFGRMLSVQSRDLRGGAFLVGMKDKPKAALSEETD 275
Db 1131 ATAKKQBIQNQNTNATTEBKQVALNQVDQELATA---INNINQADTNAEVDQAAQQLGTKAI 1187
Qy 276 -----VLGMKDKPKAALSEETDVLDAKAIKASYSYSGAALNAGKRNA----- 318
Db 1188 NAIQPNIVKKPAALAQINQHNKLAELABINATPDATNDEKNAAIN--TINODROQAIESIK 1245
Qy 319 ---GLPLD---VATDGLKAVRSLSVATSLSLTGNGLALAGVAGVSKLQKMA--TKNITD 368
Db 1246 QANTNAEVDQAAATVAENNIDAVQVDVVKKQAARDKIT-----AEVAKRIEAVKQTPNATD 1300
Qy 369 SATKAASVQSLNLSVGVGFAGWTTAGLATDPAVKKA-ESFIQDKVKSTASSTTSYV--- 424
Db 1301 EEKQAAVQINQL-----KQQAINQINQNTNDQVDTTINQAVNAIDNV 1344
Qy 425 -ADQTVK-----LAKTVKMSGEAISSTGASLSRSTVNNLHRSAPEADIEEGISAFSR 477
Db 1345 EAEVVIKPKAIAIDIEKAVKEKQQQIDN-----SLDSTDNEKEVASQALAKEKEKALAAIDQ 1400
Qy 478 SETPFQLRR 486
Db 1401 AQTNSQVNO 1409

RESULT 11
US-09-815-242-12913
; Sequence 12913, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 12913
 LENGTH: 2186
 TYPE: PRF
 ORGANISM: Staphylococcus aureus
 US-09-815-242-12913

Query Match 5.4%; Score 129.5; DB 10; Length 2186;
 Best Local Similarity 19.3%; Pred. No. 0.1;
 Matches 106; Conservative 91; Mismatches 201; Indels 151; Gaps 22;

QY 20 TADSL-----ASSVRSVSTTSCRDQAITDYKHHVFAHRSVIGSPDERDALAH 75
 DB 930 TAANTINLANTADVEQVNTAINIQGQAITPATK-----VKTDKAKA----- 972
 QY 76 EGDALVETRANLTSGETPATIAETFAKAEKFDRLATTASAFENTPPAAASVQYMQ 135
 DB 973 --IDKSAETQHNTTFNNND--ATLEEQAAQQLDQAVATAKONIN--ADTNEVA 1023
 QY 136 PAINKGDMLATPLKPLTPLISGALSGAMDQVTK---MMDRAGDLHYLSTSPDKLHA 191
 DB 1024 QANDQGNIVIQPAT-----QVKDTRVNVVDKAREATININATTGATREE 1071
 QY 192 MAVSVKHSFALGRQVNDGIAVQTSALNVRT-----VLAPALASRPVQGA--- 240
 DB 1072 KQBAIRVNTLKRRLALTDIGV--TSTAMVNSIRDDAVNQIGAVQPHVTKQTATGLNDL 1130
 QY 241 -----VDFGVSTAGGLVANNFGDRMLSVOSRDLRGAF 275
 DB 1131 AYAKKEINONTNATTEEKOVALNQVQELATA--INNINQADTNAEVAQOQLGTKAI 1187
 QY 276 -----VLGMKEKPKAALSEETDMLDAYKAIKASYSGALNAGKMA----- 318
 DB 1188 NAIQPNIVKKPALAQINQNYNAKLAENATPDAITNEKMAAIN--TLNDROQALIESIK 1245
 QY 319 -----GLPLD-----VATDGLAVRSLSVATSITKNGLALAGVAVSLQKMA--TKNITD 368
 DB 1246 QANTNAEVDQATVAENINIVAVQDVVKQQAARDKIT---AEVAKRIEAVQTPVATD 1300
 QY 369 SAKRAVSOGLSNLVGSVGFAGMTTAGLATDPAVKKA--ESFIQDKVSTASSTTSYV--- 424
 DB 1301 EEEQAAQNOINQ-----KDAQINQINQNTDQVDTTINQAVNALDNV 1344
 QY 425 -ADQTVK-----LAKTVKDMGSAISSTGASLSRTVNNLRHSAPADIEEGISAFSR 477
 DB 1345 EAEVVIKPKAIADIEKAVKEKQOQIDN-----SLDSTNEKEVASQALAKEKEKALAIIDQ 1400
 QY 478 SETPFQLR 486
 DB 1401 AQTNQVNVQ 1409

RESULT 12
 US-09-815-242-12713
 Sequence 12713, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 12713
 LENGTH: 2344
 TYPE: PRF
 ORGANISM: Staphylococcus aureus
 US-09-815-242-12713

Query Match 5.4%; Score 129; DB 10; Length 2344;
 Best Local Similarity 17.7%; Pred. No. 0.12;
 Matches 86; Conservative 97; Mismatches 222; Indels 82; Gaps 12;

QY 16 ESFRTASDASLASSVRSVST-TSCRDQAITDYKHHVFAHRSVIGSPDERDALAH 74
 DB 1421 DSGSTSTSYSTSGSESTSTSTSLDSTSIKSTSGSGSTSTASLSGSESDQSIS 1480
 QY 75 NEQIDLVETRANRLYSEGT-----PATIAETFAKAEKFDRLATTASAFENTPPAA 127
 DB 1481 TSTSEKSESTSTSLDSTSTNSGASSTSTLLNSASASESDSSTSLDSTASMQSS 1540
 QY 128 ASVLQVQWPAINKGDMLATPLKPLTPLISGALSGAMDQV----- 167
 DB 1541 ESDSQSTSTSLNSQSTSTIR--MSTIAESVSESTSESGSTSESTSTSLSDSQ 1599
 QY 168 -TQMDRAGDLHYLSTSPDKLHDANAVSVKHSFALGRQVNDGIAVQTSALNVRTV 226
 DB 1600 STSRSTASGASASTSTSDSRETSTASSTSMRTSTLDSQ--SMSLSTSTSVSD--- 1653
 QY 227 LAPALASRPVQGAUFGVSTAGGLVANNFGDRMLSVOSRDLRGAFVLGKMDKPKA 286
 DB 1654 -STSLDSVSDSTSTSTSTSGMSASISLSDSTSTSTASVW-----SA 1699
 QY 287 ALSEETDMLDAYKAIKASYSGALNAGKRMAGLPDVAITDGLKAVRSIYATSLTKNGL 346
 DB 1700 STDSQMSSESVNDSSEBSNESDS--KSMG--STSVSDG-----SLSVSTSLK--- 1749
 QY 347 ALAGYAGVSKLQKATKNITDSATRAAVSOLSNLVGSVGFAGMTTAGLATDPAVKAE 406
 DB 1750 -----SESVSEISLSGQSMDSVST-----DSSLSVSTGQSSSE 1787
 QY 407 SFIQDKVSTASSTSYVADQTVKLAITYKDMGSAISSTGASLSRTVNNLRHSAPAD 466
 DB 1788 SVSESDSLDSKSTSGSTSTSTSLSTSTSLSGSESVSESSLSDISMSDSTSTSDSD 1847
 QY 467 IEEGGS 473
 DB 1848 SLGGSIS 1854

RESULT 13
 US-09-815-242-5816
 Sequence 5816, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard

```

; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5816
; LENGTH: 2478
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5816

```

```

Query Match      5.3%; Score 127.5; DB 10; Length 2478;
Best Local Similarity 19.1%; Pred. No. 0.17;
Matches 106; Conservative 70; Mismatches 235; Indels 143; Gaps 18;

Qy 4 NQSAQPPGVAMESFRAS-----DASLASSSVRSVSTTSCRDLOAITDYLKHHVFAA 56
Db 706 NASLQDEKDVANDKIGKTIETKAIDDAATNAQVEAIKTKAINDINOTTPATTAKAAAL 765

Qy 57 HRFSVIGSPDERDAAL---AHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLA 113
Db 766 EEFDEVVQAQIDQAPLNDPTTNEEVAEAEI---RINAAKVSGVKAIEATTTAQDLERVK 821

Qy 114 TTASSAFENTPFAAASVLQYMQPAINKGDMWLATPLKPLTPLISGALSGAMDQVGTMMDR 173
Db 822 NEEISKIEN---ITDSTQTKMDAYNEVKQAATARKAQAATVSNATN---EEV----- 867

Qy 174 ARGDLHYLSTSPDKLHDAMAVSVKHSHPALGRQVDMGIAVQTFPSALNVVTVLAPALAS 233
Db 868 AEADAADVAAQKQGLHDIQVVKSKQEVADTKSKVLDKINAIOQT----- 910

Qy 234 RPSVQGVAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFLVGMKDKPEKKAALSEETD 293
Db 911 QAKVKPAADTEVENA-----YNTKQEIQNSN-----ASTTEEQAAAYTE--- 950

Qy 294 WLDAYKAIKASYSYGAALNAGKRMAGLPLDVAT---DGLKAVRSLSVATSITKNGLALAG 350
Db 951 -LDTKKQEARNLDAANTNS-----DVTAKDINSIAINQVQAATTKSKDAKAEIA 1000

Qy 351 GYAGVSKL-----QKMATKNTIDTSATKAASVQLSNLVSQVGFAGWTT-----A 394
Db 1001 QKASERKTAIEAMNDSTTEEQAAKDKVDQAVVTANADIDNAAANNVDNNAKTNEATIA 1060

Qy 395 GLATDPAVK-KAESFIQDKVKS-----TASSTTSYVAD--- 426
Db 1061 AITPDANVKAQIAIDKVAQETAIQNGNSTTEEQAAKAAQVQTEKTTADDAIDAHAH 1120

Qy 427 -----QTVKLAKTVKMSGEAISSTGASLRSSTVNNLRHSAPE-----A 465
Db 1121 TNAEVEAKKAALAKIEAIQPATTTKDKNAKEAIATKANERKTAIQAOTDITAEELIARANA 1180

Qy 466 DIEGGISAFGRSE 479
Db 1181 DVDNAVTOANSNIE 1194

RESULT 14

```

```

US-09-815-242-12967
; Sequence 12967, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12967
; LENGTH: 2478
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12967

```

```

Query Match      5.3%; Score 127.5; DB 10; Length 2478;
Best Local Similarity 19.1%; Pred. No. 0.17;
Matches 106; Conservative 70; Mismatches 235; Indels 143; Gaps 18;

Qy 4 NQSAQPPGVAMESFRAS-----DASLASSSVRSVSTTSCRDLOAITDYLKHHVFAA 56
Db 706 NASLQDEKDVANDKIGKTIETKAIDDAATNAQVEAIKTKAINDINOTTPATTAKAAAL 765

Qy 57 HRFSVIGSPDERDAAL---AHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLA 113
Db 766 EEFDEVVQAQIDQAPLNDPTTNEEVAEAEI---RINAAKVSGVKAIEATTTAQDLERVK 821

Qy 114 TTASSAFENTPFAAASVLQYMQPAINKGDMWLATPLKPLTPLISGALSGAMDQVGTMMDR 173
Db 822 NEEISKIEN---ITDSTQTKMDAYNEVKQAATARKAQAATVSNATN---EEV----- 867

Qy 174 ARGDLHYLSTSPDKLHDAMAVSVKHSHPALGRQVDMGIAVQTFPSALNVVTVLAPALAS 233
Db 868 AEADAADVAAQKQGLHDIQVVKSKQEVADTKSKVLDKINAIOQT----- 910

Qy 234 RPSVQGVAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFLVGMKDKPEKKAALSEETD 293
Db 911 QAKVKPAADTEVENA-----YNTKQEIQNSN-----ASTTEEQAAAYTE--- 950

Qy 294 WLDAYKAIKASYSYGAALNAGKRMAGLPLDVAT---DGLKAVRSLSVATSITKNGLALAG 350
Db 951 -LDTKKQEARNLDAANTNS-----DVTAKDINSIAINQVQAATTKSKDAKAEIA 1000

Qy 351 GYAGVSKL-----QKMATKNTIDTSATKAASVQLSNLVSQVGFAGWTT-----A 394
Db 1001 QKASERKTAIEAMNDSTTEEQAAKDKVDQAVVTANADIDNAAANNVDNNAKTNEATIA 1060

Qy 395 GLATDPAVK-KAESFIQDKVKS-----TASSTTSYVAD--- 426

```

Db 1061 AITPDANVKAQAIDKQAOETAIDGNGSTTEEKAQVQTEKTADAIDAHA 1120
 QY 427 -----QTVLAKTVKMSGSAISSTGASLSTVNNLRHSAPB-----A 465
 Db 1121 TNAEVEAKKAIAKIEAIGPATTTKONAKAIAITKANERTALAQODITAEIIAANA 1180
 QY 466 DIEGGISAPSRSE 479
 Db 1181 DVDAVTOANSNIE 1194

RESULT 15

US-09-137-531-15
 ; Sequence 15, Application US/09137531
 ; Patent No. US20020048816A1
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Expression of surface layer proteins
 ; NUMBER OF SEQUENCES: 25
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/137,531
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/682,517
 ; FILING DATE:
 ; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1222 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-137-531-15

Query Match 5.2%; Score 125.5; DB 10; Length 1222;
 Best Local Similarity 20.3%; Pred. No. 0.091;
 Matches 117; Conservative 73; Mismatches 219; Indels 167; Gaps 25;

QY 3 INSAQOPGVAME-----SFTASDASLASSSVRSYSTT 37
 Db 675 VMSGKKVTGTSIKKATYITNTGANDIKVNDQVISPNRSYTVYEATLSST---GTVIT 731
 QY 38 SCRDLOAITTYLKHVFAARFSYISGSPDERDALAHNEQIDALVEF----- 84
 Db 732 PAKNLEVTSYDGKTAVKVIATGIAVNTDGDVAFATAKEATATPTATNEVPNSYTGVAQ 791
 QY 85 -----RANRLYSEGETPATIAETPAKAEKPRRLATTASAFENTPFAASVLOV--- 133
 Db 792 FNTADSGSNSNIWFAKGNVKTAGVSGKTYK--FGANGNEVFGAAMEAL-LTQYATE 848
 QY 134 -----MOPAINKGDMLATPLKPLTPL--ISGALS-----GAMD--- 164
 Db 849 GQKVTISYVNDGPTVTFKVISAVNSSTEAIKPAFTTPAAPTGGALTLTPRAGGLVDLT 908
 QY 165 QVGTAKMDRARGDLHISTSPDKLHDMAMVSK-RHSPALGROVDMGIAVQTPSALNV 223
 Db 909 ATNTLGISLADALNVSATTV---DTATVSLDSANNSLSLTIVETGANTGVFAF--- 960
 QY 224 RTVLAAPALASRPVQAVDVGSTAGLVAN-----AGFGDRLMSVQSRDQLRG--- 273
 Db 961 -TVQAGTSS-----LTAGLTIVTYADAKAAAGVAENITASVTLKKTGTGITS 1007
 QY 274 -AFVLGKDKPEKPAALSEETDMLDAYKAIRASYSQAL--NAGKMAGLPLDVATDGL 329
 Db 1008 DFTFGVL--PSAATAEYTSKSIADYTFATGEGFTLNIDNAGQVINLA-----GK 1058
 QY 330 KAVRSLSVSAITSLKNGALAGVAGVSKLOKMTKNITD-----SATKAASGOLSNLV- 382

Db 1059 KGAQVADAINATFAGTAVSG-----DKVVIKSAITGVSGSEVEVTFSSVNOVLNAV 1111
 QY 383 -GSVGVFAGW-TTAGLATDPAVKKASFIODKVKST-----ASSTSY 423
 Db 1112 NGKDQVAVAGTAATKAFITTTALSVGEKVVIDGVEYTRAVAGTAPANTFVESNANTLAS 1171
 QY 424 VADQTVLAKTVKMSGE--AISSTGA--SLRSTV 454
 Db 1172 VADQANLAAITDITPLNADKFTASATGATITLTSTV 1207

Search completed: January 2, 2003, 15:22:17
 Job time: 19.5087 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 2, 2003, 15:19:06 ; Search time 8.49126 Seconds
(without alignments)
1084.696 Million cell updates/sec

Title: US-09-825-414-7

Perfect score: 2412
Sequence: 1 MHNRVQOPVATDSEFR.....IEEGTAASPSRIFRPMRS 486

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCR_NEW_PUB pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*
4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*
7: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2412	100.0	486	10 US-09-825-414-7	Sequence 7, Appl1
2	1869.5	77.5	487	10 US-09-825-414-66	Sequence 66, Appl1
3	146	6.1	1463	9 US-09-971-536-69	Sequence 69, Appl1
4	143	5.9	600	10 US-09-287-849-22	Sequence 22, Appl1
5	133.5	5.5	729	10 US-09-287-849-2	Sequence 2, Appl1
6	131.5	5.5	2086	10 US-09-815-242-5639	Sequence 5639, Ap
7	131.5	5.5	5795	10 US-09-815-242-12610	Sequence 12610, A
8	129.5	5.4	2344	10 US-09-815-242-12713	Sequence 12713, A
9	129	5.3	2434	10 US-09-815-242-5835	Sequence 5835, Ap
10	129	5.3	6281	10 US-09-815-242-12996	Sequence 12996, A
11	127.5	5.3	596	10 US-09-287-849-26	Sequence 26, Appl1
12	127	5.3	2122	9 US-09-813-214A-9	Sequence 9, Appl1
13	126.5	5.2	1026	10 US-09-379-931-7	Sequence 7, Appl1
14	125	5.2	714	10 US-09-841-786-4	Sequence 4, Appl1
15	125	5.2	3241	10 US-09-841-786-1	Sequence 1, Appl1
16	124.5	5.2	1744	9 US-10-108-605-25	Sequence 25, Appl1
17	122	5.1	5215	9 US-09-860-846-2	Sequence 2, Appl1
18	122	5.1	5215	10 US-09-861-289-2	Sequence 2, Appl1
19	120.5	5.0	1879	9 US-09-971-536-70	Sequence 70, Appl1

20	118	4.9	2285	10 US-09-932-183A-2	Sequence 2, Appl1
21	116.5	4.8	573	10 US-09-828-574-13	Sequence 13, Appl1
22	116	4.8	2478	10 US-09-815-242-5816	Sequence 5816, Ap
23	116	4.8	2478	10 US-09-815-242-12967	Sequence 12967, A
24	116	4.8	2993	9 US-09-738-626-6239	Sequence 6239, Ap
25	115	4.8	342	10 US-09-815-242-11043	Sequence 11043, A
26	115	4.8	1832	9 US-10-014-717-4	Sequence 4, Appl1
27	114	4.7	2368	10 US-09-815-242-5635	Sequence 5635, Ap
28	114	4.7	2368	10 US-09-815-242-12389	Sequence 12389, A
29	114	4.7	7257	9 US-10-014-717-5	Sequence 5, Appl1
30	113.5	4.7	650	10 US-09-815-242-13341	Sequence 13341, A
31	113	4.7	596	9 US-10-063-547-100	Sequence 100, App
32	113	4.7	596	12 US-10-006-867-100	Sequence 100, App
33	113	4.7	596	12 US-10-052-586-310	Sequence 310, App
34	113	4.7	1488	9 US-09-738-626-5495	Sequence 5495, Ap
35	112	4.6	2076	10 US-09-815-242-5815	Sequence 5815, Ap
36	112	4.6	2186	10 US-09-815-242-12913	Sequence 12913, A
37	112	4.6	4999	9 US-09-576-059-14	Sequence 14, Appl1
38	111.5	4.6	1289	9 US-09-712-363-259	Sequence 259, App
39	110	4.6	647	10 US-09-841-132-436	Sequence 436, App
40	110	4.6	792	9 US-10-025-380-1127	Sequence 1127, App
41	109.5	4.5	1186	10 US-09-826-752-8	Sequence 8, Appl1
42	108	4.5	856	10 US-09-287-849-12	Sequence 12, Appl1
43	107.5	4.5	1165	12 US-10-016-768-10	Sequence 10, Appl1
44	107.5	4.5	1221	12 US-10-016-768-11	Sequence 11, Appl1
45	107	4.4	1054	10 US-09-798-042-87	Sequence 87, Appl1

ALIGNMENTS

RESULT 1

US-09-825-414-7
Sequence 7, Application US/09825414
Patent No. US20020083489A1
GENERAL INFORMATION:
APPLICANT: Collmer, Alan
APPLICANT: Alfano, James R.
APPLICANT: Charkowski, Amy O.
TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
FILE REFERENCE: 19603/3243
CURRENT APPLICATION NUMBER: US/09/825,414
CURRENT FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/194,160
PRIOR FILING DATE: 2000-04-03
PRIOR APPLICATION NUMBER: 60/224,604
PRIOR FILING DATE: 2000-08-11
PRIOR APPLICATION NUMBER: 60/249,548
PRIOR FILING DATE: 2000-11-17
NUMBER OF SEQ ID NOS: 91
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 486
TYPE: PRT
ORGANISM: Pseudomonas syringae
US-09-825-414-7

Query Match 100.0%; Score 2412; DB 10; Length 486;
Best Local Similarity 100.0%; Pred. No. 4.3e-176;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MHNRVQOPVATDSEFRASDASLSSSVRSVSSQOOREINAIADYLDHVFRAHKL P 60
DB	1	MHNRVQOPVATDSEFRASDASLSSSVRSVSSQOOREINAIADYLDHVFRAHKL P 60
QY	61	PADSADQAAVDVNAQITLIEFRASRLHFEGETPATIADTFKAKELDRLATTTSCAL 120
DB	61	PADSADQAAVDVNAQITLIEFRASRLHFEGETPATIADTFKAKELDRLATTTSCAL 120
QY	121	RATFPAAASLLQYQAPINKGDWLPAPLKPPLTPLISGALSGAMDQVGTMMDRATGDLHY 180
DB	121	RATFPAAASLLQYQAPINKGDWLPAPLKPPLTPLISGALSGAMDQVGTMMDRATGDLHY 180

QY 181 LSASPDRLHDMAASVKRHSPLARQVLDGTGAVQVTSARNVAVRTVLAPALASRPVQGA 240
DB 181 LSASPDRLHDMAASVKRHSPLARQVLDGTGAVQVTSARNVAVRTVLAPALASRPVQGA 240
QY 241 VDLGVSAGGAAANAGFGRNLLSVQSRDHQGGALVGLGDKPKAQLSEENDWLEAYKA 300
DB 241 VDLGVSAGGAAANAGFGRNLLSVQSRDHQGGALVGLGDKPKAQLSEENDWLEAYKA 300
QY 301 IKSASYSGAALNAGKRWAGLPDMATDAMGAVRSLSVASSLTQNGLALAGGFAGVGLQ 360
DB 301 IKSASYSGAALNAGKRWAGLPDMATDAMGAVRSLSVASSLTQNGLALAGGFAGVGLQ 360
QY 361 IKSASYSGAALNAGKRWAGLPDMATDAMGAVRSLSVASSLTQNGLALAGGFAGVGLQ 360
DB 361 IKSASYSGAALNAGKRWAGLPDMATDAMGAVRSLSVASSLTQNGLALAGGFAGVGLQ 360
QY 421 TGYVADQTVKLAKTVKMGGEAITHGTGASLRNTVNNLRQRPAREADIEEGGTAA 480
DB 421 TGYVADQTVKLAKTVKMGGEAITHGTGASLRNTVNNLRQRPAREADIEEGGTAA 480
QY 481 FRPMRS 486
DB 481 FRPMRS 486

RESULT 2
US-09-825-414-66
; Sequence 66, Application US/09825414
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; CURRENT FILING DATE: 2001-04-03
; PRIOR FILING DATE: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR FILING DATE: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR FILING DATE: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 66
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Pseudomonas syringae pv. tomato
US-09-825-414-66

Query Match 77.5%; Score 1869.5; DB 10; Length 487;
Best Local Similarity 78.4%; Pred. No. 8e-135;
Matches 381; Conservative 33; Mismatches 71; Indels 1; Gaps 1;
QY 1 MHINRRVQPPVTADSPRTASDASLASSSVRSVSSDOOREINAIADYLTIDHVFAAHKL 60
DB 1 MHINQSAQPPGVAMESFRTASDASLASSSVRSVSTTSCRDQLAITDYLKHVFAAHRFS 60
QY 61 PADSADGQAAVDVHNAQITAIETRASTRLEHFEGETPATIATFAKAEKLDRLATTTGAL 120
DB 61 VIGSPDERDAALAHNEQIDALVETRANLYSEGETPATIATFAKAEKFDRLATTTASSAF 120
QY 121 RATPAMASLLQYWPALNKGDWLPAPLPLTPLISGALSGAMQVGTAKMDRATGDLHY 180
DB 121 ENTFFAASVLYQWPALNKGDWLPAPLPLTPLISGALSGAMQVGTAKMDRATGDLHY 180
QY 181 LSASPDRLHDMAASVKRHSPLARQVLDGTGAVQVTSARNVAVRTVLAPALASRPVQGA 240
DB 181 LSTSPDKLHDMAVSVKRHSPLARQVLDGTGAVQVTSARNVAVRTVLAPALASRPVQGA 240

QY 241 VDLGVSAGGAAANAGFGRNLLSVQSRDHQGGALVGLGDKPKAQLSEENDWLEAYKA 300
DB 241 VDLGVSAGGAAANAGFGRNLLSVQSRDHQGGALVGLGDKPKAQLSEENDWLEAYKA 300
QY 301 IKSASYSGAALNAGKRWAGLPDMATDAMGAVRSLSVASSLTQNGLALAGGFAGVGLQ 360
DB 301 IKSASYSGAALNAGKRWAGLPDMATDAMGAVRSLSVASSLTQNGLALAGGFAGVGLQ 360
QY 361 MATKNIITDPATKAAVQSOLTNLAGSAAVFAAGWTTAALTTPPAVKKAESFIQDTVKSTAS 420
DB 361 MATKNIITDPATKAAVQSOLTNLAGSAAVFAAGWTTAALTTPPAVKKAESFIQDTVKSTAS 420
QY 421 TGYVADQTVKLAKTVKMGGEAITHGTGASLRNTVNNLRQRPAREADIEEGGTAA 479
DB 421 TGYVADQTVKLAKTVKMGGEAITHGTGASLRNTVNNLRQRPAREADIEEGGTAA 479
QY 480 FRPMRS 485
DB 481 FRPMRS 486

RESULT 3
US-09-971-536-69
; Sequence 69, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Blokeberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christensson, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
; TITLE OF INVENTION: Using Them
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR FILING DATE: 2001-10-02
; PRIOR FILING DATE: 2000-08-08
; PRIOR FILING DATE: 2000-08-08
; PRIOR FILING DATE: 2000-11-28
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1463
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-69

Query Match 6.1%; Score 146; DB 9; Length 1463;
Best Local Similarity 22.5%; Pred. No. 0.0049;
Matches 112; Conservative 72; Mismatches 219; Indels 94; Gaps 20;
QY 14 ATDSFRTASDASLASSSVR---SVSSD--QOREINAIADYL-----TDHVFAAHKLPPA 62
DB 996 ASEAKASSNASTAATAAAGVFAASDASQAATAAGADVASSAATAANSNAAAAASAT 1055
QY 63 DSADGQAAVDVHNAQITAIETRASTRLEHFEGETPATIATFAKAEKLDRLATTT 116
DB 1056 KAGDSKAAAGFSSAASAAASAKGAEAVASEAASAAASDDSVASSAASAAAGFKAASAA 1115
QY 117 SGALRATPFAASLLQYWPALNKGDWLPAPLPLTPLISGALSGAMQVGTAKMDRATG 176
DB 1116 EGAASSAASAAAS-----SAAAGQ-----TRGGASSSSASEAG-----QASTA 1152
QY 177 DLHYLSAPDRLDHMAASVKRHSPLARQVLDGTGAVQVTSARNVAVRTVLAPALASRPA 236

Db 1153 TSYVASASSASE--AGSTAHQAGSSASEA--TGHASSATSQASA-----ASSASRRYP 1203
QY 237 VOGAVDLGVSMAGGLAANAGFGRNLLSVQSRDHORGALVGLKDKPEKQQLSEENDMLE 296
Db 1204 SDGSIQDVSIASSAASSTAS-----SASASAGSEASTASASASHASEQASIASSEDEVIS 1257
QY 297 AYKAIKASISGGAALNKG---BMAGLPLDMATDMAGAVSLVSASSLTONGLALAGGF 352
Db 1258 SSASAVSASSAASASAAKAGNSGAGIYSHASASAAASAKSASAESQASASASAAA----- 1312
QY 353 AGVAGKLOEMATKNTDTPATAVSQLTNLAGSAVFGMTTAAITTPDPAV-----KKA 405
Db 1313 -----SDSVASSAASALSDDAKASSAADVASSATTAAISATSISLADQSGATGTA 1363
QY 406 ESFIQDTVKSTASTTGV-ADQT-----VTLAKTVKMGGEALTHTGASLRTVNN 456
Db 1364 GSHI---LPSIGETTSIPSGCPTPTQTKPTQTKPTQTKPTQACQT--TQTC-SLPQDHA 1418
QY 457 LRQRPAREADIEEGTA 473
Db 1419 GRHMLPQTGDDESGTS 1435

RESULT 4
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tr1-fusion
US-09-287-849-22

Query Match 5.3%; Score 143; DB 10; Length 600;
Best Local Similarity 23.0%; Pred. No. 0.0024;
Matches 117; Conservative 58; Mismatches 165; Indels 168; Gaps 25;
QY 20 TASDASLASSVSVSSDQOREINAIADYLDHVFPAHKLPDASDQGA----- 69
Db 88 TAGQAEITLAAQVRV-----AAAAYETAYGLT---VPPVIAENRLEMLILATNL 134
QY 70 -----AVDVHNAQITALETRASRLHFEGETPATIADTFPAKAEKLDRLATTSGAL--- 120
Db 135 LCGWTPAIANVAEYGMNQDAAMFGYAAATATATATLLPFEAEPM--TSAGGLLEQ 192
QY 121 -----RATPFAMAS-LIQYQOPAIKNGDWLPAKLPLTPLISGALSGAMDVG-----T 168

Db 193 AAIVEASDTPAANQLMNNVPOALQO---LAOPTQCTTP--SSKLGGLMKTVSPHNSPIS 247
QY 169 KMMDRATGDLHYLSASPDRLHDMAASVKHSPSLARQVLDTGAVQVYASAPNAVITVLA 228
Db 248 NMTSMANNHNS-MTNGSVSTNTLSSWLGKFAFAAAQO-----AYQT-AAQNGVR----- 295
QY 229 PALASRPVQAVDLGVSMAGGLAANAGFGRNLLSVQSRDHORGALVGLKDKPEKQQL 288
Db 296 -AMSSLSISLSSGLG---GVAANIG---RAASVGS-----L 326
QY 289 SEENDMLEAKIKASAS-----YSGALNKGKMAGLPLDMATDMAGAVSLVSASS 340
Db 327 SVPOAMAAANQAVTPAARALPLTSLTSAERGPQMLGILPVG----- 369
QY 341 LTONGLALAGFPAGV-----GKL-----QEMATKNTIDP- 369
Db 370 --QMGARAGGLSGVLRVPRPVPMPHSPAGLDDPDVAINTCNGVYVAAALNATDPG 427
QY 370 -----ATKAASVQLTNLAGSAVFGMTTAAITTPDPAVKASFIQDTVKSTASTTGY 423
Db 428 AAQFNASPVAGSYLRNFFLAAPPQRAAMAAQL---QAVGMAQYI-GLVESVAGSCNMY 483
QY 424 VADQTVLAKTVKMGGEALTHTGASLR 451
Db 484 E-----LMTINYQFGDVDAH-GAMIR 503

RESULT 5
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tr1-fusion
US-09-287-849-2

Query Match 5.5%; Score 133.5; DB 10; Length 729;
Best Local Similarity 21.0%; Pred. No. 0.017;
Matches 116; Conservative 70; Mismatches 178; Indels 189; Gaps 25;
QY 20 TASDASLASSVSVSSDQOREINAIADYLDHVFPAHKLPDASDQGA----- 69
Db 221 TAGQAEITLAAQVRV-----AAAAYETAYGLT---VPPVIAENRLEMLILATNL 267
QY 70 -----AVDVHNAQITALETRASRLHFEGETPATIADTFPAKAEKLDRLATTSGAL--- 120

Db 268 LQONTAIVNEAEYGEWMAQDAAMFCYAAATATATATLTPFEAPEM--TSAGGLLEQ 325
 Qy 121 -----RATPFAMAS--LLQYMOAIPAINKGWLPAPKLPPLISGALSGAMDQVG-----T 168
 Db 326 AAAVEEASDTAAANQLMNVFOALQO---LAQPTGGTTP--SSKLGGLWKTVPSPHSPI 380
 Qy 169 KMDRATGDLHYLSASPRDLHDMAASVKRHSPLARQVLDTGAVQVYSARNVTRVLA 228
 Db 381 NNVSMANNHMS--MTNSGVSMNTLTSSMLKGFAPAAARQ-----AVQT-AAONGVR----- 428
 Qy 229 PALASRPVQGVADLVGSMAGLAANAGFGRNLLSVQSRDHQGGALVLGDKPEKPAQL 288
 Db 429 -AMSSLSGSSGLG---GVAANLG---RAASVGS-----L 459
 Qy 289 SEENDWLEAYKAISAS-----YSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASS 340
 Db 460 SVPQAWAANAQAVTPAARALPLTSLSAARGPGQMLGGLPVG-----502
 Qy 341 LTQGLALAGGFAGVKL-----QEMATKNITDPA-----TKAAVSQ 377
 Db 503 --QMGARAGGGLSGVLRVPPRPVYMPHSPAAAGDIAPPALSQDRFADFPALPLDPSAMVAQ 560
 Qy 378 L-----TNLAGSAAFAGTWTAAALTTDP-----AVKKAESFIQDTVKSTA 417
 Db 561 VPQVNVNTKLVYNAVAG--TGIVDPNGVVLNHHVIAGATDINAFSVSGQTYG 617
 Qy 418 SSTTGYVADQTVKLAK-----TVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEE 469
 Db 618 VDVVGVDRTQDVAVLQLRGAGGLPSAAGGIVGAVGEPVVMGNS-----661
 Qy 470 GGTAAASPEIIFR 482
 Db 662 GGQGTTPRAVPCR 674

RESULT 6

US-09-815-242-5639
 ; Sequence 5639, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011a
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5639
 ; LENGTH: 2086
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus

US-09-815-242-5639

Query Match 5.5%; Score 131.5; DB 10; Length 2086;
 Best Local Similarity 21.6%; Pred. No. 0.1;
 Matches 124; Conservative 70; Mismatches 226; Indels 153; Gaps 26;
 Qy 3 INRVQ--QPPVTATDSFRTASDASLASSVRSV-----SSDQOREI 42
 Db 185 INNAVQANKRTATIKNGTAMPTNLAGSGTTTIPVTVTYNDGSTEEVOESIFTKADREL 244
 Qy 43 NATADYLTDRHFAAHKLPPADSADGQAADVHNAO---ITALIETRASRLHFEGETATI 99
 Db 245 ITAKNHLDDPVSIDGKKPGTITQYNNR---IHNAQQOINTAKTEAQ--QVINNERATQQV 300
 Qy 100 ADTFAKAELDRLATTTSGALRATPPFAMASLLQ-----YMQPAINKGDLWLA 146
 Db 301 SDALTKV---RAAQTKINEAKA-----LLQNKEDNSQLVTSKNNLQSSVNO----- 343
 Qy 147 PLKPLTFLISGALSGAMDQVGTMMDBRATGDLHYLSASPRDLHDMAASVKRHSPLARQ 206
 Db 344 -----VPSITGMTQQSIDNYNAKKREAE-----EITAAQRVIDNGDATAQQ 385
 Qy 207 VLDTGVAVQVYSARN-----AVRTVLAPALASRPVQGVADLVGSMAGLAAAGF 257
 Db 386 ISENTTAQRNDLITNQISOATNLAAVESVKQSA--NSLDGANGNLQTAINDKSGTLASQNF 444
 Qy 258 GNRLLSVQSRDHQGGALVLGDKPEKPAQLSEENDWLEAYKAIKSASYSGAALNAGKRM 317
 Db 445 -----LDADQEKRNAYNQAVSNAE--TILNKQTGNTAKTAVEQALNN---VNSAKHA 492
 Qy 318 AGLPLDMATDAMGAVRSLVSASSITQNG-----LALAGGFAGVGLQEMATKNITDPA-- 371
 Db 493 LNTQNLNNAKQAAITAINGASDLNKKQKDALKAQANGAQRVSNQAQDV--QRNATELNTAM 551
 Qy 372 ---KAAVSQLTN--LAGSAAVFAAGWTTA--ALTTPAVKKAESFIQ--DTVKSTASTTGYV 424
 Db 552 GTLKHAIDAKTNTLASSKYVNAOSTKQNAIYT--KVTNAEHIISGTFTVTVTTPSEVTA 609
 Qy 425 -----ADQTVKLAK-----TVKDMGGEA-----ITHT 446
 Db 610 NQVNSAKQELNGDBERLRVAKQNAITAIDALTQNLTPQKAKLKEQVQGNARLEDIQTQVTN 669
 Qy 447 GASLRNTVNNLRQRPAREADIE--EGGTAASPS 477
 Db 670 GOALNNAMKGLRDSIANETTVKASQNYTDASPN 702

RESULT 7

US-09-815-242-12610
 ; Sequence 12610, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011a
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23

```

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12610
; LENGTH: 5795
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-09-815-242-12610

Query Match          5.5%; Score 131.5; DB 10; Length 5795;
Best Local Similarity 21.6%; Pred. No. 0.42;
Matches 124; Conservative 70; Mismatches 226; Indels 153; Gaps 26;

QY 3 INRRVQ-QPVTATDSFRTASDASLASSVRSV-----SSDQREI 42
DB 2462 INNAVQANKRTATIKNGTAMPTMLAGSSTTIPVTYNDGSTEEVQESIFTKADREL 2521
QY 43 NAIDVLTDFHFAHKLPPADSADGQAAVDVHNAQ---ITLITRASRLHFEGETPATI 99
DB 2522 ITAKNHLDDPVSTDGKKGITTYNNNA---IHNAQQQINTAKTEAQ-QVINERATPOQV 2577
QY 100 ADTFPAKLEKLDRLATTTSGALRATPFAMASLQ-----YMQPAINKGWLPA 146
DB 2578 SDALTYV---RAQOTINEKA-----LLQNKENSQLVTSKNNLQSSVNG----- 2620
QY 147 PLKPLTPLISGALSGAMDQVGTCKMDRATGDLHYLASPDRLHDMAASVYKRSPSLARQ 206
DB 2621 -----VSTTGMTQOSIDNYNAKKEAEFT-----ETTAQRVIDNDGATAAQ 2662
QY 207 VLDTGVAVQVTSAN-----AVRTYLA PALASRPVAGADVLSMAGLAANAAGF 257
DB 2663 ISDENTTAQRNDLTNQISQATNLAAVESVQSA-NSLDGAMGNQOTAIINDKSGTLASQNF 2721
QY 258 GNRLLSVQSRDHQRCALVLGLKDEKPPAQLSEENDWLEAYKAISKASYSALNAGKRM 317
DB 2722 -----LDNDEQKRNAYNOVNSNAE--TILNKOTGRTACTAVYQALNN--VNSAKHA 2769
QY 318 AGPLDMATDMAGVRSLSVSSSLTONG---LALAGFAGVGLQEMATKNITDPAT-- 371
DB 2770 LNTQNLNNAKQAALITAINGASDLNQKQKDALKAQANGAQVSNADQV-QNATLNTAM 2828
QY 372 ---KAASQULTN-LAGSAAVFAGWTTA-ALTDPAYKKAESFIQ-DIVKSTASTTGYV 424
DB 2829 GTLKHALADTKNTVLASSKRYNADSTKQNAVTT-KVTNAEHIIISGTPVTTPSEVTA 2866
QY 425 -----ADQTVKLAK-----TVKDMGGEA-----ITHT 446
DB 2887 NQVNSAKQELNGDERLRAVAKQONATATDALQTQLTPQAKLKEQVQANRLDITQVQTN 2946
QY 447 GASLRNTVNNLRQRPAREADIE--EGGTAASPS 477
DB 2947 GQALNNAMKGLRDISANETTVKASQNTYDASPN 2979

RESULT 8
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in

```

```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12713
; LENGTH: 2344
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match          5.4%; Score 129.5; DB 10; Length 2344;
Best Local Similarity 17.6%; Pred. No. 0.17;
Matches 86; Conservative 107; Mismatches 204; Indels 91; Gaps 18;

QY 8 QQPVATATDSFRTASDASLASSSVSSVSDQOREINAIADVLTDFHFAHKLPPADSADG 67
DB 764 QQSQSVTSKADQSASTSTSGSIMTSTAASKTSVS--LDSVASAKSLSTSEBN-- 819
QY 68 QAAVDVHNAQITLIERASRLHFEGE-TPATTAIDF---AAAEKLDRLATTTSGALRA 122
DB 820 ---SVSSSTSTSLVNSQSVSSMSGVSKSTSLSDPISNSSSTKESBSVSTSTSLRT 875
QY 123 TPFAMASLQYMQPAINKGWLPAPLKPLTPLISGALSGAMDQVGTCKMDRATGDLHYLS 182
DB 876 S-----TSLSDSVMSWTSGSLK--SQSLSTSD-----S 904
QY 183 ASBDR-LHDMAASVYKRSPSLARQVLDTGVAVQVTSARNAVETVLA PALASRPVAGAV 241
DB 905 ASISQSVSDSTNSI-----STSELSSEGSTSESIISINSISVSAISGKLESQSTSI 959
QY 242 DLGVSAAGGLAANAAGGNRLLSVQSRDHQRCALVLGLKDEKPPAQLSEENDWLEAYKA 301
DB 960 SLSTSDSKSMSTSESLSD--STSTSDSVSGSLVAG---SQSVSTSTSDSMSTSEMI 1011
QY 302 K-SASVSGAALNKGKMAAGLPLMATPDAMG---AVRSIVSASSLTQNGLALAGFAGVG 356
DB 1012 SDKMSSTSGSLAASDSKMSVSSSMSTSGSTSESLSDSISTSDSDSKSLSTSGSG-- 1069
QY 357 KQEMATKNITDPATKAAYVQLTNLGSAAVFAGWTTAALTDPAYKKAESFIQDTYKST 416
DB 1070 -----STSTSTSSSVRNSQSTSGSMSTSGSDSTSI-----SFDSTSDSK 1115
QY 417 ASSTGTVADQTVKLAKTVK-DWGGEALITTGASLRN-----TVNNLRQRPAREADIE 469
DB 1116 SASSTA-----SSESISQSVSTSTSGSVSTSTSLSTNSERTSTSMDSSTSLSTSESD 1170
QY 470 GGTAAASPS 477
DB 1171 DSTSTSDS 1178

RESULT 9
US-09-815-242-5835
; Sequence 5835, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert

```

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12996
; LENGTH: 6281
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-09-815-242-12996

      Query Match          5.3%; Score 129; DB 10; Length 6281;
      Best Local Similarity 19.7%; Pred. No. 0.72;
      Matches 112; Conservative 89; Mismatches 237; Indels 130; Gaps 23;

Qy 3 INRRVQPPVTATDSFTASDASLASSVSRSVSDQOREINAIADYLTDHVFAAHKLPPA 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 871 LNQKQDALKTVNNAQGVSDANNVQHTATELSAMTAKAAIAD--KERTKASGNYNA 928
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 63 DSADGQA-----AVDVHNAQIT-----AL1ETRASRLHFGEETPATIADTFAK 105
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 929 DQEKRQAYDSKVNTNAENIISCTPNATLTVDNVNSAASQVNAAKTALNGDNNLRVAKEHAN 988

Qy 106 AEKLDRLIATTTSGALRATPFAMASLLQYMPAINGKDWLPAPLPLPLTISGALSGAMDQ 165
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 989 -----NTIDGIAQLNNAQKAKLKEQVQSATTLDG--VQTVKNSSTLTNTAMKGLRDS 1038

Qy 166 VGTKMMDRATGDHLHYLSASPRDLHD-----AMAASVKRHS-----PSLARQVLD----- 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1039 IANEAT1KA--GQNTYDASPNRNEYDSAVTAAKAIINQTSNPTMFWNTTITQVTSQVTTK 1096

Qy 210 -----TGVAVQYTSARNNAVRTVLAPALASRPVQGVADLVGSVMAG---GLAANAGFGN 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1097 EQALNGARNLAQAATTAKNNLNLTSINNAQKDALTRSIDGATTVAGVNOQTAKATELNN 1156

Qy 260 RLLSVQSRDHQRGALVLGLKDKPEKPAQLSEENDWLEAYKA1KASASYSGAALNAGK----- 315
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1157 AMHSLON-----GIND---ETQTKQTKYLDAPESKGA--YDOAVNNAKAITT 1200

```


Db 564 LTVGNSTLNDGLSVKNTSNKQIQVGADGTFITDINSKPGAGI--ENTRIITRDGIGF 621
Qy 304 ASYSGAALNAGK-RMAGLPLDMATDAMGAVRSLVSASSLTQGLALAGGFAVGKL---Q 359
Db 622 ANNTG-SLDANKPRLTPTGINAGGKELTNVQSAINPAT---NG-----GQLDFMN 667
Qy 360 EMATKNTIDPATKAASVOLTNLASAAVAFAGTWTAAALTTDPVAKKARSFIQDTVKSTASS 419
Db 668 RLSTANTEKSGSAATIKDLYNLQVPLTFAGDTGPNVTK---KLGEILKVKGGKTTADD 723
Qy 420 TT---GVVADQ-----TVKLAKTVKMGGEAITHTGASLNTV---NLLRQRPAREADI 467
Db 724 LTKNNIGVADVSTNSLTVKLAKTLDLSDAVNTKLTASDKVTVDGNNNTAKLQNGDLTF 783
Qy 468 EGGTAASPS 477
Db 784 SKONTGATPA 793
RESULT 13
US-09-379-931-7
; Sequence 7, Application US/09379931
; Patent No. US2002009792A1
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. US2002009792A1ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULI
; FILE REFERENCE: 08106/002003
; CURRENT APPLICATION NUMBER: US/09/379,931
; CURRENT FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/614,377
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: US 08/194,290
; PRIOR FILING DATE: 1994-02-09
; PRIOR APPLICATION NUMBER: US 07/895,367
; PRIOR FILING DATE: 1992-06-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-379-931-7

Query Match 5.2%; Score 126.5; DB 10; Length 1026;
Best Local Similarity 22.2%; Pred. No. 0.092;
Matches 113; Conservative 56; Mismatches 177; Indels 163; Gaps 23;
Qy 12 VTATDSFRTASDASLASSSVRSVSSDQOREINAIADY----- 48
Db 206 VSGIGGYATATAAMINDLSDGALSTDNAAGVNLFTAYPSSGVSGSTLTTGTDTLTGTA 265
Qy 49 -----LTHVFAAHLPPADGADGAAVDVHN-----AQITAL-----IETRASRLHF 91
Db 266 NNDTFVAGEVAGAAATLVGDTLGGAGTDLVNWQAAVATLPTGVTISGIE-----MNV 321
Qy 92 EGEPATITADTFAKAEKLDRLATTTSGA-----LRATPFAMASLLQYMQPAINK 140
Db 322 TSGAAITL-NTSSGVGTGLTALNTNTSGAAQVITAGAGQNLITATTAQAANNVAVDGRAN- 379
Qy 141 GDWLPAPILKPLTPLISGALSAMQVGTQKMDRATDGLHYLSASPDRLHDMAASVKKHS 200
Db 380 -----VTVASTGVTSGT-TTVGAN--SAASG-----TVSVSVANS 412
Qy 201 PSLARQVLDTGAVQT--YSARNVARTVLAAPALSRPAVQAVDLGVSMAGLAAAGFG 258
Db 413 TTTTGAIATVGTGATVTAQTAGNANTTLTQADVTVTGNSSTTAVTVTQTAATAGATVA 472
Qy 259 NRLLSVQSRDHQRCALVLGLKDKPKAQLENDWLEAYKAIKASYSYSGAALNAGKMA 318
Db 473 GRV-----NGAVTI-----TDSAAASATTAGK--- 494

Qy 319 GLPLDMATDAMGAV-RSLVSASSLTQGLALAGGFAVGKLOEMATKNTITDTPATKAASVQ 377
Db 495 -----IATVTLGFGAATIDSSALTTVNLSGTGTSIGIR-----GALT--ATPTANTL 541
Qy 378 LTNLAGSAAVAFAGTWTAAALTTDPVAKKARSFIQD---TVK---STASSTTGYVADQTVKL 431
Db 542 TLNVG-----LTTTGAITDSEAAADDGFTTINIAGSTASST---IASLVAAD 586
Qy 432 AKTVKMGGEAIT---HTGASLREN-TVNN 456
Db 587 ATTINISGDARVITTSHTAAALGTIIVTN 615
RESULT 14
US-09-841-786-4
; Sequence 4, Application US/09841786
; Patent No. US20020054883A1
; GENERAL INFORMATION:
; APPLICANT: NAGARAJA, T. G.
; APPLICANT: STEWART, GEORGE C.
; APPLICANT: NARAYANAN, SANJEEV K.
; APPLICANT: CHENGAPPA, M. M.
; TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN
; TITLE OF INVENTION: VACCINE AND PREPARATION THEREOF
; FILE REFERENCE: 30296
; CURRENT APPLICATION NUMBER: US/09/841,786
; CURRENT FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 09/558,257
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Fusobacterium necrophorum
US-09-841-786-4

Query Match 5.2%; Score 125; DB 10; Length 714;
Best Local Similarity 20.3%; Pred. No. 0.072;
Matches 113; Conservative 91; Mismatches 220; Indels 132; Gaps 25;
Qy 14 ATDSFRTASDASLASSSVRSVSSDQOREINAIADYLTDHVFAAHLPPADSADGQ----- 68
Db 73 AVNIQSKTSALVKNSDIRNAN---KPKVNALSG--GTQVAAGAGLEAVKESGGQGSYL 127
Qy 69 ---AAVDVHNAOITALITETRASRLHPEGETPATIADTFAKAEKLDRLAT-----TT 116
Db 128 LGTSASINLVNNEVSASSENNT-----VAGESEQKMDVDVTAYQADTVTGALNLQAGKS 183
Qy 117 SGALRATPFAMASLLQYMQPAINKGDWL---PAPLKLPLTELISGALSAMQVGTQKMDR 173
Db 184 NGTVGAT-VTVAKLNNKVNASISGRVTVNVNRADAKAL--LATTQVTAATVTTGTTISSGA 240
Qy 174 ATGDLHYLSASPDRLHDMAASVKKRHS-----LARQVLDTGAVAQVYSARNVRTV 226
Db 241 GLGN-YQGVASVKNIDNVEASVDKSGIEGANBINIAKDVKGSSDLAKHEYQ----- 291
Qy 227 LAPALSRPAVQAVDLGVSMAGG-----LAANAGFNGNRLLSYQSRDHQRCG-- 273
Db 292 ---ALLNGDKKYLEDRGINTTNGYTYTKEQLEKAKKKEGAVIVNAALSVAAGTDKSAGV 348
Qy 274 ALVLGLKDKPEKQOLSEEN-----DWLEAYKAIKSASYSYSGAALNAGKMA----- 318
Db 349 AIAVNTVKNKPKAELSGNKEAGEDKIHAKHNVNVEAKSSTVWVNAASGLAISKDAFSGMG 408
Qy 319 -GLPLDMATDAMGAV-RSLVSASSLTQ-----GLALAGGFG-----VG----- 356
Db 409 SGAWQDLSNDTIKAVDKRGISADSLNVNANNISILGNVNAVAGTIAGSLSTAVGAAPANNLH 468
Qy 357 -KLOEMATKNTIDP-----ATKAARV-----SQTNLIA--GSAAVFAGTWTAAALTTDPAVKK 404
Db 469 NKTSALITGVKNPFFSGKNTKVNVOALNDSHITNVSAGGAASIKQAGIGMVSVNRSDE 528

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:16:50 ; Search time 29.3692 Seconds

(without alignments)
3341.397 Million cell updates/sec

Title: US-09-825-414-7

Perfect score: 2412
Sequence: 1 MHNNRKYQPPYATDSFRT.....IEEGTAAPESEIIPRPMRS 486

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2412	100.0	486	2 Q9JP34	Q9JP34 pseudomonas
2	1653	68.5	484	2 Q8RP03	Q8RP03 pseudomonas
3	490	20.3	518	16 Q8XQF0	Q8XQF0 ralsstonia s
4	153	6.3	1545	16 Q9DD01	Q9DD01 streptomyc
5	151.5	6.3	516	16 Q8SS82	Q8SS82 ralsstonia s
6	150.5	6.2	2055	2 Q8S472	Q8S472 abiotrophia s
7	149	6.2	1953	16 Q9BHJ2	Q9BHJ2 rhizobium l
8	148	6.1	1203	5 Q9NSK0	Q9NSK0 caenorhabdi
9	147.5	6.1	1363	2 Q9FSK8	Q9FSK8 rhodospacer
10	147.5	6.1	1822	2 Q07290	Q07290 streptococ
11	147	6.1	1156	16 Q9Z5A4	Q9Z5A4 streptomyc
12	147	6.1	6048	2 Q93H87	Q93H87 streptomyc
13	145.5	6.0	973	16 Q8XDQ4	Q8XDQ4 escherichia
14	144.5	6.0	2124	16 Q98M03	Q98M03 rhizobium l
15	144.5	6.0	2232	5 P91365	P91365 caenorhabdi
16	144	6.0	515	16 Q8XT14	Q8XT14 ralsstonia s

17	143.5	5.9	1713	3 Q8TGE1	Q8TGE1 saccharomy
18	142.5	5.9	949	4 Q9Y4G6	Q9Y4G6 homo sapien
19	142	5.9	2016	5 Q9BRT0	Q9BRT0 plectreury
20	141.5	5.9	481	2 Q93GT1	Q93GT1 campylobact
21	141.5	5.9	3381	2 Q9KX33	Q9KX33 streptococ
22	139	5.8	1794	9 Q9T1A7	Q9T1A7 bacterioph
23	138.5	5.7	2155	16 Q8UFP9	Q8UFP9 agrobacteri
24	138	5.7	506	2 Q53B34	Q53B34 salmonella
25	138	5.7	825	17 Q8RTJ9	Q8RTJ9 methanosarc
26	137.5	5.7	5636	5 Q9N9M2	Q9N9M2 leishmania
27	137	5.7	506	2 Q53B36	Q53B36 salmonella
28	136.5	5.7	491	2 Q8RTY4	Q8RTY4 campylobact
29	136.5	5.7	653	2 Q9X6N1	Q9X6N1 rhizobium l
30	136	5.6	597	16 Q9S2K2	Q9S2K2 streptomyc
31	136	5.6	4776	16 Q97P71	Q97P71 streptococ
32	135.5	5.6	1731	16 Q8U8W4	Q8U8W4 agrobacteri
33	135.5	5.6	2479	16 Q9A988	Q9A988 caulobacter
34	135	5.6	718	16 Q9JZD0	Q9JZD0 neisseria m
35	135	5.6	1128	17 Q52009	Q52009 halobacteri
36	134	5.6	492	2 Q93GT3	Q93GT3 campylobact
37	134	5.6	2271	16 Q99QY4	Q99QY4 staphylococ
38	133.5	5.5	466	16 Q8BLA8	Q8BLA8 rhizobium l
39	133.5	5.5	5644	2 Q93NX8	Q93NX8 streptomyc
40	133	5.5	506	2 Q53837	Q53837 salmonella
41	133	5.5	1208	16 Q9PFI9	Q9PFI9 xyloella fas
42	132.5	5.5	487	2 Q93GT2	Q93GT2 campylobact
43	132.5	5.5	1336	16 Q9ACV2	Q9ACV2 streptomyc
44	132	5.5	540	9 Q9G0H8	Q9G0H8 roseophag
45	132	5.5	1762	2 Q52546	Q52546 amycolatops

ALIGNMENTS

RESULT 1

ID Q9JP34 PRELIMINARY; PRT; 486 AA.

AC Q9JP34;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 50.7 kDa protein.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
NCBI_Taxid=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=98422476; PubMed=9748456;
RA Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
RA Collier A.;
RT "The Pseudomonas syringae pv. tomato HrpW protein has domains similar
RT to harpins and peptidate lyases and can elicit the plant hypersensitive
RT response and bind to peptate.";
RL J. Bacteriol. 180:5211-5217(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=20243785; PubMed=10781092;
RA Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,
RA Petricki-Owleja T., van Dijk K., Collier A.;
RT "The Pseudomonas syringae Hrp pathogenicity island has a tripartite
RT mosaic structure composed of a cluster of type III secretion genes
RT bounded by exchangeable effector and conserved effector loci that
RT contribute to parasitic fitness and pathogenicity in plants.";
RN Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX Ramos A.R., Rehm A.H., Collier A.R.;
RT "Pseudomonas syringae pv. tomato DC3000 hrpL through hrpC";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

[illegible]

```

Db 98 SESSNSVSGFMAASIRKQSIGISNSPPGSGPHPTDPCGTGCTIDIPHMPIADQ 157
Qy 49 LTHVFAHKLPPADSADGQAAVAVHNAQTALITRASHLHEGETPATIADTFAKAEK 108
Db 158 LRD-IFRPH-----FDG-----TNDQREILLQGAERLQMGEPATVAVALKAGAN 204
Qy 109 LDRLATTSALATPPMAASLQYMPAINKGDMPLKPLTPLISGALSGMDQVGT 168
Db 205 RDLRLAQTVGFVRSVPFGISARLPFDVQALTAFA-----AKTTARVAGATVAGSGTADAFGC 260
Qy 169 KMDRATGDLHYLSASPPDRLDHMAASVKKHSPSLARQVLDGVAQVTSARNAVTVLA 228
Db 261 TLGKATSNQWMLAASDHLPEPMAQAKVQPSLGRLLAASVSLAFQYSLRNVRIGVA 320
Qy 229 P-----ALASRPVQAVDLGVSMAGLAANAFGNRLLSVQSRDHQCGALVLGKDEP 284
Db 321 PLATHALGAARAA--VDSNIAVGVPAAGAAATMAQHNETHRRGAYVLG----- 372
Qy 285 KAQLSENDVLEAVKAIKASYSYGALNAGKRAGLPL-DMAIDMGAVALSVASSLTQ 343
Db 373 -----RTDWDQFKQKOSTWTDPLVAGAKRTAKLVLADLTETTLAATRSLEPATWLK 426
Qy 344 NGLAGGFAGVGLQEMATKNITD-----PATRAVSQLTNLAGSAVPAFGMTTALT 399
Db 427 MGLLAGGFAGVGLTQTAAGAAATEAGYTEAAVAVRAVSTVLSAPVYAAWTTADVMA 486
Qy 400 PAVKAESEPIQ 410
Db 487 PAIDAAAGHIQ 497

RESULT 4
Q9RDO1 PRELIMINARY; PRT; 1545 AA.
ID Q9RDO1
AC Q9RDO1
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative secreted protein.
GN SC02383 OR SC4A7.11.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleeser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleeser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleeser J., Lathe L., Murphy L., Oliver K., O'Neill S.,
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,

```

```

RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL133423; CAB62715.1;
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00652; Ricin_B_lectin; 3.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
SQ SEQUENCE 1545 AA; 161520 MW; 81EF325143593AEA CRC64;

Query Match 6.3%; Score 153; DB 16; Length 1545;
Best Local Similarity 22.6%; Pred. No. 5.5;
Matches 113; Conservative 66; Mismatches 202; Indels 118; Gaps 20;

Qy 6 RVQGPVTATDSTRRTASDASLASSSVRSVSPQQRINAI-----ADYLTDHV 53
Db 745 RAETADTTNTERKKGAEARDKAKLRDADWDQKANAARAKAKAEVAQASASEDNA 804
Qy 54 FAHKLPPADSADGQAAVAVHNAQTALITRASHLHEGETPATIADTFAKAEKLDRL 112
Db 805 QEARK--AANDADA-AADAEPTAAVAAKSEADKATTAADADADAATRAEAAKRSADSD 861
Qy 113 AT-----TSGALRAATPFMAASLQYMPAINKGDMPLKPLTPLISGALSGMDQV 167
Db 862 AAQAALTLDAAVRTATVSAADAIRKASKTAATNA-----RTA 898
Qy 168 TKMDPATGDLHYLSASPPDRLDHMAASVKKHSPSLARQVLDGVAQV-----YSARNV 223
Db 899 VELADDA--EQH-----AADAKKEADAKAEAVT-ALAGANESTGYVTTAQAADVAGNSA 951
Qy 224 RTVLAPALASRPVQ--GAVDLGVSMAGLAANAFGNRLLSVQSRDHQCGALVLGKOK 282
Db 952 AQVAAPA--NDALIQSGSPVTTDSAGLAVLTGQSSKTTAEQ--QQVVA----- 996
Qy 283 EPRAQLSEEN-----DWLEAYK-AIKSASYSYGALNAGKRAGLPLDMATDA 328
Db 997 EAAVQAQAESEAQAQSVANASGDSKAAVYTLAALGVAAADANNSKAEALGVSAEAASTA 1056
Qy 329 MCAVRSLV-----SASSLTONGLALAGFPAGVGLQEMATKNITDPAT 371
Db 1057 TQAQSLTRTTIADTQATDAAADSAAGRAEGHABDARDSDAALDAEAAASADTAE 1116
Qy 372 KAASQLTNLAGSAVFAAGTTTALTDPVAVKAESEFIDTYKSTASTTGYVADQTVKL 431
Db 1117 QAAED-----ARDADAHA-AETAAAEEPAKDAQRYAESAQQA-----AEQAEKE 1160
Qy 432 AKTVKDMGGEALITHTGASL 450
Db 1161 ANAEQIDKGTVDVDTGAPI 1179

RESULT 5
Q8XS2 PRELIMINARY; PRT; 516 AA.
ID Q8XS2
AC Q8XS2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Probable flagellar hook-length control protein.
GN FLIK OR RSP0395 OR R800823.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plaemid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823652;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattoiico L.,

```

RA Chandler M., Choisine N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex I.,
 RA Siguier P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 DR EMBL; AL646078; CADI7546.1; --
 DR InterPro; IPR000104; Antifreeze1.
 DR PRINTS; PR00308; ANTIREEZE1.
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 516 AA; 50024 MW; 8969D64D4F43ECF4 CRC64;

Query Match 6.3%; Score 151.5; DB 16; Length 516;
 Best Local Similarity 24.0%; Pred. No. 1.5;
 Matches 121; Conservative 46; Mismatches 208; Indels 129; Gaps 19;

Qy 10 PPVATDFTASDASLASSVSSVSSQREINAIADYLDHVFPAHKLPPASADQQA 69
 Db 5 PNTAATDLGAMSSAGTAADKSQAASSADTFGNLLSRLNQE--AARQTDASAGSRA 62
 Qy 70 AVDVHNAQITAIETFRASRLHFEGETPATIADTFKAEEKLRLATTSTGALRATPFAMAS 129
 Db 63 A-----ARR-----ADATKDAKASAAAGN-----82
 Qy 130 LLOYMOPAINKDWLPAPLKPPLTPLISGALSGAMDQVGTQMDRATGDLHYLSASPDRLH 189
 Db 83 -DQDARQADRNADAAFA-----TAAASATQDAQASK-----AASSDDTQ 120
 Qy 190 DAMAASVRRHSPSLARQVLDGTGVAQVTSARNVAVTVLAPALASRPVQGVAV--DLGVSM 247
 Db 121 DKTAQAQGVDA-----AQLAAQIEAARQMQQAAP--VSGPTGVATADDAQAOL 171
 Qy 248 AGGLAANAGFNGRLSVQSRDHORGALVLGLKKEPKAQLSEENDWLEAYKAIKSASYS 307
 Db 172 AAGKAGATTDAALAAALSAANAAGKLT-----DPAATVAQP-----QAQAADALK 218
 Qy 308 GAALNAGRMAGLPL-DNATDAMG-----NVRSLVSASSLTQ-----NGLAL 348
 Db 219 AAAGKEAQAQAKPTPLPDASTVAQALARTGTDTNTPPAVRHLGSSASSAQAAGSERGSA 278
 Qy 349 AGGFAGVGLKQEMATKNITDPATKAASVQLTN-----LAGSAVFAVGTWTAALTDPVKK 404
 Db 279 HRTEAGQGTQQAATAETAEGRNADTSNPTQADAGAGAKFDGVLTTRA--NGDAVAT 336
 Qy 405 AESFTQDTVKSTAGSTGGVADQTVKLAKTVKMGGEAITHTGASL-----RNTVNNL 457
 Db 337 APTF---IVGAGTAGTASAAAVTPQAQHTLPTFGDAAMPHTMASQLAYMQVHRQSSAEL 393
 Qy 458 RQRPAR-----EADIEEGGTAAS 475
 Db 394 QLSPAELGLGLHVKLEIDNGAVNAS 417

RESULT 6
 O85472 PRELIMINARY; PRT; 2055 AA.
 ID O85472;
 AC O85472;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Extracellular matrix binding protein (Fragment).
 GN EMB.
 OS Abiotrophia defectiva.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Aerococcaceae; Abiotrophia.
 OX NCBI_TaxID=46125;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NVS-47;
 RX MEDLINE=99081722; PubMed=9864195;
 RA Manganelli R., van de Rijn I.;
 RT "Characterization of emb, a gene encoding the major adhesin of
 RT Streptococcus defectivus";

RL Infect. Immun. 67:50-56(1999).
 DR EMBL; AF067776; AAD03320.1; --
 DR TIGRFAMs; TIGR01168; YSIRK_signal; 1.
 FT NON TER 2055 2055
 SQ SEQUENCE 2055 AA; 215642 MW; 9699C11DDDE93E2FD CRC64;

Query Match 6.2%; Score 150.5; DB 2; Length 2055;
 Best Local Similarity 22.2%; Pred. No. 11;
 Matches 126; Conservative 73; Mismatches 246; Indels 123; Gaps 21;

Qy 3 INRRVQOPPVTA-----TDSFTASDASLA-----SSSVSVSSDQOREINAIADYLDH 52
 Db 726 INEISQPDUTREKQAFMDQVTRTADANAKVASANNQAVTSARDQGLNANNLPT-- 783
 Qy 53 VFAAHLPPA-----DSADQAAVVDVHNAQITAIETFRASRLHFEGETPATIA----- 100
 Db 784 --PAKYPEALGHVRQAADAKRQAIRDNANLTAEEQADALRQVDAQAATAEAALNQHN 841
 Qy 101 DTFKAEEKLD-RLATTTSGALRATPFAMASLLQY---MQPAINKGDLWLPAPLK-PLTPLI 155
 Db 842 ATLAADSDGVKAINDINPQPSKPAANQALEQVAAAKRQAINNNNLTDEEKAQAIQV 901
 Qy 156 SGALSGAMDQV-----GTMKMDRATGDLHYLSASPDRLHDAMAASVKRHSPLARQVL 208
 Db 902 DQALANAKTVQRAANDNGVNOA-KTAGTTAINNINPQGTQKQAIAIEAASQAKRDEL 960
 Qy 209 DTGVAVQVTSARNVAVTVLAPALASRPVQ-----GAVDLGVSMAGGL----- 251
 Db 961 QGRNDLTTEERNALADLTAKAQAADAVNQARNNTGVAGAKDNGVAQIQGINPTAVVKP 1020
 Qy 252 -AANAGFNGRLSVQSRDHORGALVLGLKKEPKAQLSEENDWLEAYKAIKSASVSGA 309
 Db 1021 DARNA-----IDQAARDKEAFQANTKLTDEEKAAIKKVQDAARDAKAAIDRAGSNGD 1074
 Qy 310 ALNA-----GKEMAGLPLDMDATMGAVRSVLSASSLTQNGLALA 349
 Db 1075 VNNAVNGKAAIQAIKALDSDQSADKTAQAIONADAKKA--AITANNALTQEEKA-- 1130
 Qy 350 GGFAGVGLKQEMATKNITDPATKAASV-----QLTNLAGSAAVFAGWTT 393
 Db 1131 -----AAIKQVEDEAQAQAADVASRSKADVDRAKDGLOKISDPAVQPPKLN 1179
 Qy 394 AALTDPVAVKKAESFTQTVKST---ASSTTGVADQTVKLAKTVKMGGEA-ITHTGAS 449
 Db 1180 AIAAVDQAATDKKAVINNDTTLTQEEKEAIRKVDEEAKARQAINDATSNADVAAKQAO 1239
 Qy 450 LRNTVNNLRORP---AREADIEEGGTA 473
 Db 1240 GTQAINNVPTPAKNAKAAVEQAADA 1267

RESULT 7
 Q98HJ2 PRELIMINARY; PRT; 1953 AA.
 ID Q98HJ2;
 AC Q98HJ2;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein ml12848.
 GN ML12848.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFP303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;

RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti.";
 RT DNA Ref. 7:331-338(2000).
 DR EMBL; APO03000; BAB49874.1; -;
 DR InterPro; IPR000911; Ribosomal_L11.
 DR PROSITE; PS00359; RIBOSOMAL_L11; UNKNOWN_1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 1953 AA; 184557 MW; 892BFA8B687B35E2 CRC64;
 SQ

Query Match 6.2%; Score 149; DB 16; Length 1953;
 Best Local Similarity 23.1%; Pred. No. 12; Indels 140; Gaps 23;
 Matches 123; Conservative 53; Mismatches 216; Indels 140; Gaps 23;

QY 2 HINRRVQPPVATDSE-----RTASDASLASSSVRSVSDQREINAIADYLTDHVF 54
 DB HANSTLADSSATGDSVAIGPTSTATSSAIAAGSNANAGAN-----ASAIIGTSSVASAL 884
 QY 55 AAHKLPPADSADGQAADV-HNAQITLITRASRLHFEGETPATIADTPAKAEKLDRLA 113
 DB DATAMGFLSKASQGFSTAVGANANATLSTALGQNALASGVQATLQKQANASASDALA 944
 QY 114 -----TTTSGALRA-----TFPMASLLQYMOFPAINKGDL-----PAFLKPLT 152
 DB LGANSTRAGNAGVALSGSVTAAGVTPNAINGTTFYAFQGINPASTVSGAPAEKTLT 1004
 QY 153 PLISGALSG-----ANDQVGTMMMDRAT-GDLHYLSASPDRLHDAMA-- 193
 DB 1005 NLAAGRSSGSTDAVNSQLFATNQAVDAIGTNNISTGGIGIKYFPAN-STLADSSATG 1063
 QY 194 ---ASVGRHSPLARQVLDTG-----VAVQTSANNAVTVL----- 227
 DB 1064 TDSVAIGPTSTATSSAIAAGSNANAGANASAIIGTSSVASALDATAMGFLSKASQGFST 1123
 QY 228 -----APLALSRPANOAGVNDGV-SMAGGLANAGFGNRL-LSVQSR-----DHQRG 273
 DB 1124 AVGANANATLSTALGQNALASGVQATLQKQANAGASDALAGANSTAGNAGVALGS 1183
 QY 274 ALVGLKDEPKKAQLSEENDMLEAYKAIKSAS-----YSGAALNAGKRMAGLPIDMATDA 328
 DB 1184 GSTTAAVAGTPNVI---NCTTYAFQGINPASTVSGAPAEKTLITLAAGRISGSSTDA 1240
 QY 329 MG-----AVRSLVSASSLTONGALAGFAGVGKLGEMATKNTTDPATKAASQLTNLAG 383
 DB 1241 VNSQLFATNQAVDAIGTNNI--LGGSVTNIGV-----VNNIAG 1279
 QY 384 SAAVFAQWTA-----ALITDPAVKASFIQDTYKSTASSTTGVADOT 428
 DB 1280 DTS--TAYTDANGIGIRYARTNEAGLQDTSFAG-----GLGSTAVGYQASAT 1325

RESULT 8

Q9NSKO PRELIMINARY; PRT; 1203 AA.
 AC Q9NSKO;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical 120.1 kda protein.
 GN H43B16.1.
 OS Caenorhabditis elegans.
 OC Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoidea; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OC NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=9069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;
 RA Miller N., Carter T.;
 RT "The sequence of C. elegans cosmid H43B16.";
 RL Submitted (MAR-1999) to the EMBL/Genbank/DBD databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBD databases.
 DR EMBL; AC006669; AAF39909.1; -;
 DR InterPro; IPR000082; SEA_domain.
 DR SMART; SM00200; SEA; 1.
 KW Hypothetical protein.
 KW SEQUENCE 1203 AA; 120104 MW; E04CADC2BA74713B CRC64;
 SQ

Query Match 6.1%; Score 148; DB 5; Length 1203;
 Best Local Similarity 20.3%; Pred. No. 7.3; Indels 72; Gaps 19;
 Matches 105; Conservative 83; Mismatches 258; Indels 72; Gaps 19;

QY 12 VTATDSFRTASDASLASSSVRSVSDQREINAIADYLTDHVFPAHKLPPADSADGQA 71
 DB 454 VTSTPTI-TSTQAMASSSSENPSTIQ-----AISSYTTVASSTIAPISSESSIAS 507
 QY 72 DVHNAQITLITRASRLHFEGETPATIADTPAKAEKLDRLATT-----TSG 118
 DB 508 SSAPSGSTVITVSTTAVSSGGSTGFTTIGGSSFGSGSTIQTGTSSFAPIPSTAG 567
 QY 119 ALAATPPMAAS-----LQYQWPA--INKGDLPAFLKPLPLISGALSGAMDQVGT 169
 DB 568 SSGQTPGSMSTGTGVMSSSFQPTAPFSLGTMTPGTSSIPISISTVNSGSSSTGT 627
 QY 170 MMDRATGDLHYLSASPDRLHDAMAASVKRHSPELARQVLDTG-----VAVQTSARNA 222
 DB 628 VTQAPSS---TSMGSPSGOSTGTMNTSAPFYTSSSANTGSTSGTIVTQVSSR 684
 QY 223 VRTVLAAPALASRPAGVAVDLGVSMAGLAANAGFNRLLSVQSRDHQRCALVLGLKDK 282
 DB 685 TSPV---ASSSQMTSTQCPGSGSSISGSTVNGSSS--VTTQPPASRSTA-SQSSSA 728
 QY 283 EPKAOQLSEENDMLEAYKAIKSASYSGAALNAGKRMAGLPIDMATDMGAVRSLVSASLT 342
 DB 739 QPTASSSTMGSTAGSSSPQPTASSTVPSSTGATSSGSTVSGST--MGSTQSSLPSSTMT 796
 QY 343 QNGLALAGFAGVGKLGEMAT--KNITDPATKAASQLTNLAGSAVFA----- 389
 DB 797 NTG---STGSTVNQALASSSTIGASSTEPASTANPGSSTGQIATVTTGSSSTQTN 853
 QY 390 GMTTAAITTDPAVKA---ESFIQDTVKSTA---SSTTG--VYADQTVKLAKTVKDMG-- 439
 DB 854 TSGSTGTVINPVSSTSGSSSTQPIASSTANPGSSTGTVTTTGGSSSTQTNPTGSR 913
 QY 440 GEAITHTGASLRNTVNNLRQPPAREADIEBGGTAAPS 477
 DB 914 GSTVTPSAFSSSTASSTQPIASSTANPGSSTGPT 951

RESULT 9

Q9FSK8 PRELIMINARY; PRT; 1363 AA.
 AC Q9FSK8;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Flagellar hook-associated protein 1.
 GN FLGK.
 OS Rhodospirillum rubrum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;
 OC Rhodospirillum.
 OC NCBI_Taxid=1063;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gonzalez-Pedraza B., De la Mora J., Ballado T., Camarena L.,
 Dreyfus G.;

RT	Query Match	Best Local Similarity	Matches 132;	Conservative 47;	Mismatches 193;	Indels 157;	Gaps
RT	Isolation and Complementation of a Flagellar P-ring Mutant of						
RL	Rhodobacter sphaeroides.;						
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.						
DR	EMBL: AF317649; AAG31286.1; .						
DR	InterPro: IPR001444; Flag_bb rod.						
DR	InterPro: IPR001899; Gram_pos_anchor.						
DR	Pfam: PF00460; flg_bb rod; 1.						
DR	PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN 1.						
DR	PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN 1.						
DR	SEQUENCE 1363 AA; 133154 MW; AD7BB4856117808C CRC64;						
QY	56 AHKLPAPDASGQAAVDVHN-----AQTALITETASRLHFECETPATIADTFAK 105						
Db	592 AGSLTAEPAGGTQIITDGRHIAAGALTAEEALLITEAN-----GFLPGAYVD--- 642						
QY	106 AEKLDRLATTSSGA----LEATPFANASL-----LOYMQPAINKGDWLP--APLKPLT 152						
Db	643 -----ASTLNGAGGTGFRGTGIAGAILPGERVLSLHPADVPAGSSGLLPASALPSLT 695						
QY	153 -----PLISGALSAMDQ-----VGTQMDRATGDHLY-----LSA 183						
Db	696 LEAAGGLPLVQLPAGASAAEMAQAINAFAGIEAEARTVTIEAPADGTTLTFTLTGNL 755						
QY	184 SPDRLHDMAASVKRHPSPLARQVLDTGVAVQTYGARNAVRTVLAPALASRPVQ--GAVD 242						
Db	756 SPVRISGAVAGG---RNDALA-----LAVNAVSAATGVRAEUSPDGARILLVQDGGAD 805						
QY	243 LGYSMAGLAANAAGFNRLLSVQSRDHQGGALVL--GLKDKEPKAQLSEENDWLEAYKAI 301						
Db	806 IGIV---GLRHTAGAAVTLQCTDAES PAGAPLTLSETAD---SARFTGE-----L 850						
QY	302 KSASYSYGAAALNAGKRMWAGPLDMATDAM--GAVRSLSVASSLTQN-----GLA 347						
Db	851 RLSSASGFSADLG---GVRQDAADVPMSSGGLVSRGVSAGGVQTYGYTDPAPFDGAGLS 906						
QY	348 LAGGFAGVGLQEWAT---KNITDPATKAASQLTNLAGS-----AAVFAGWTTA 394						
Db	907 ADGTFQAQSAQAQVAMTVGNRTVLTDLAAAAGVSDGAGVASAALALRAEAPARTTGGPVA 966						
QY	395 ALTTDPRAKVAESFIQDVKSTASSITGVVA-----DOT-----VKLAKTVKD 437						
Db	967 AL----PADGRSVSVSEGSYTLRMTGGVAVDGGEPGLLTAAPATNRLVIQAAGSLDG 1023						
QY	438 MG-----GEAITHGTASLRNTVNNLRQRPAREA-----DIEEGTAAAS 475						
Db	1024 AGLRIESGAAAFGLAAADAPVSTLTGQPADPALPASFDIELGCTLYS 1072						
RESULT 10							
Q07290	PRELIMINARY;						
ID	Q07290						
AC	Q07290						
DT	01-NOV-1996 (TrEMBLrel. 01, Created)						
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)						
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)						
GN	EF protein.						
DE	EF protein.						
OS	Streptococcus suis.						
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;						
OC	Streptococcaceae; Streptococcus.						
OX	NCBI_TaxID=1307;						
RN	[1]						
RC	SEQUENCE FROM N.A.						
RP	STRAIN=1890;						
RX	MEDLINE=93328288; PubMed=8335363;						
RA	Smith H.E., Reek F.H., Vecht U., Gielkens A.L.J., Smits M.A.;						
RT	"Repeats in an extracellular protein of weakly pathogenic strains of						
RT	Streptococcus suis type 2 are absent in pathogenic strains."						
RT	Infect. Immun. 61:3318-3326(1993).						


```

Db 5848 PVLRLVPAARTRARA-AAAPSSASSAERLEENPEAER-EKTVLDLVRAEVAAGLGH 5905
QY 425 ADQTVKLAKTVKMGGEAITHGTASLRNTVN 455
Db 5906 SDRTVRPEHAFQDLGFDLSL--TAVELRRLN 5934

RESULT 13
Q8XDQ4
ID O8XDQ4 PRELIMINARY; PRT; 973 AA.
AC O8XDQ4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative membrane protein of prophage CP-933X (Putative tail fiber protein).
DE
DE GN Z1918 OR ECS1650.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouzis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11259796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005333; AAC56007.1; ALT INIT.
DR InterPro; IPR00104; Antifreeze_1.
DR InterPro; IPR004089; Cmtaxis_transd.
DR InterPro; IPR005003; Phage_fiber.
DR Pfam; PF03335; Phage_fiber.
DR Pfam; PF03406; Phage_fiber_2.
DR PRINTS; PR00308; ANTIFREEZE1.
DR Complete proteome.
RW SEQUENCE 973 AA; 96316 MW; 776580D2A87E1B36 CRC64;

Query Match 6.0%; Score 145.5; DB 16; Length 973;
Best Local Similarity 22.5%; Pred. No. 7.4;
Matches 107; Conservative 52; Mismatches 232; Indels 85; Gaps 15;

QY 19 RTASDASLASSSVRSVSDQOREINAIADYLTDFVFAAHKLPADSDAQVADVHNAQI 78
Db 117 RNASVAQNTAAAKSASDASTSAREATHATD---AADSARAASTSAGQAASSAQSS 173
QY 79 TALIETRSLRHFEGETATTDTEAKELDRLATTTGALRATPFAMASILQYMOPAI 138
Db 174 SA-----GTASTKATEASKSAAAEKSSAAATSAAGAKTSETNNAVSQ--QSA 221
QY 139 NKGDWLPAPLPLTPLISGALSAMDQVGTQKMDRATGDLHLVLSASPDRLHDMAA--SV 196
Db 222 TSA-----STATTKASEFAASARDASAKSEAKSS-ETSAASSASSAASSATAAGNSA 273

```

```

QY 197 K-----RHPSLAROVLDTCGVAQVQYSA--RNAVRTVLAPALASRPVQAVDGLV 245
Db 274 KAAKTSETNAKSSETAEEQASASAAAGSKTAALASASASTSAGQASASATAAGSAESA 333
QY 246 SMAGLAANAGFGNRLLSVQSRDHQRGALVGLKDKPEKAQLSEENDWLEAYKAIKSAS 305
Db 334 SSASTATTKAGEATEQASASASASAAKTSNNAKSETSAESSKTAASASSASASAS 393
QY 306 YSGAA-----LNAGKRMAGLPDMDATDAMGAVRSIVLSASSLTQNLGLAGGAGV 358
Db 394 SASASKDEATROASAAKSSATTASTKATEAAGSATAAQAQSKSTAESAATRA----- 444
QY 359 QEWATKNITDPA-----TKAAVSQLTNLGSAVAVFAGWTTAALTDP-AVKKAES 407
Db 445 -ETAAKRAEDIASVALEDASTTKGIVQLSSATNS-----TSESLAATPKAVKAAYE 496
QY 408 F-----IQDTVK-----STASSTTGVVADQTVKLAKTVKDMGGEAITHGTS 449
Db 497 LANGKYTAQDATTAQKGIQVQLSNATNSTSEMLATPKSVKAAAYDLANGKYTAQDAT 552

RESULT 14
Q98M03
ID Q98M03 PRELIMINARY; PRT; 2124 AA.
AC Q98M03
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Kinesin-like protein.
GN MLR0796.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002995; BAB48310.1; -.
RW Complete proteome.
SQ SEQUENCE 2124 AA; 229253 MW; 3457A2B462EFED01 CRC64;

Query Match 6.0%; Score 144.5; DB 16; Length 2124;
Best Local Similarity 21.2%; Pred. No. 24;
Matches 108; Conservative 86; Mismatches 191; Indels 125; Gaps 22;

QY 19 RTASDASLASSSVRSVSDQOREINAIADYLTDFVFAAHKLPADSDAQVADVHNAQI 450
Db 405 RTTGMVSLGGAARTLSEFEASLNGIERTLAER-----GQALISEFQTRA 450
QY 70 -ADVHNAQITALTETASRLHFE-GETPATIADTFK-----AEKLDK----- 111
Db 451 EALDTGTQKLNAALEARQINETLVEREIAHTFAESKDTLAAMIDQGTQIGADWAD 510
QY 112 LATTSGALRATPFAMASLLQYMQPAINKGDWLPAPLPLTPLISGALSAMDQVGTQRM 171
Db 511 IVTSTSMLEARSDFAGRMEAAHVVSR--FSDIQLADARVG-IEEAVENTHSRK-- 565
QY 172 DRATGDLHLVLSASPDRLHDMAAASVKSHSPSLARQVLDTVGAVQTSYARNVTRVLAP 231
Db 566 -----LSESDRMAAQAQDLKFAES--RDGIDAAVTNQVKLAEG-RSLIARAL 613
QY 232 ASRPVQAVDGLVSMAGGLAANAGFGNRLLSVQSRDHQRGALVGLKDKPEKAQLSEE 291
Db 614 EE-----DLRKVNESRAIDAISLGSHLERLSE-----GRNRLSL-----ALNED 652

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:13:55 ; Search time 34.4645 Seconds
(without alignments)
1879.027 Million cell updates/sec

Title: US-09-825-414-7

Sequence: 1 MHINRRVQQPPTATDSFRT.....IEEGTAASPSEIPFRPMRS 486

Scoring table: BLOSUM62

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database :

A_Genseq.101002.*

1:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/genseq/genseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2412	100.0	466	22	AAE12573	Pseudomonas syringae
2	2412	100.0	486	22	AAE67677	Amino acid sequencer
3	1869.5	77.5	487	22	AAE12603	Pseudomonas syringae
4	146	6.1	1463	23	AAE20110	Lactobacillus rhamnosus
5	144.5	6.0	1822	13	AAE27745	Extracellular factor
6	143	5.9	600	20	AAU32068	Mycobacterium tuberculosis
7	143	5.9	600	23	AAU74597	Antigenic fusion protein
8	142.5	5.9	1134	22	AAU78810	Human protein SEQ
9	142.5	5.9	1144	22	AAU39293	Human polypeptide
10	142.5	5.9	1165	22	AAU39312	Human polypeptide

11	139.5	5.8	1177	22	AAAM9794	Human protein SEQ
12	139.5	5.8	1177	22	AAAM1019	Human polypeptide
13	139.5	5.8	1177	22	AAAM1098	Human polypeptide
14	135	5.6	723	21	AAE15477	Neisseria meningitidis
15	133.5	5.5	729	21	AAE17572	Mycobacterium species
16	132	5.5	1721	19	AAAM52847	A. mediterranei rib
17	132	5.5	1787	19	ABBA9791	Listeria monocytogen
18	131.5	5.5	2086	21	AAAG4143	Staphylococcus aure
19	131.5	5.5	3596	22	AAAG7407	Bordetella pertussis
20	131.5	5.5	5795	22	AAU37017	Staphylococcus aure
21	130.5	5.4	1995	18	AAAG2567	Staphylococcus aure
22	130.5	5.4	1996	18	AAAG23717	Platenolide synthas
23	130.5	5.4	3647	11	AAK05041	Platenolide synthas
24	129.5	5.4	1201	22	ABBE1629	Flamentous haemag
25	129.5	5.3	2344	22	AAU37120	Drosophila melanog
26	129	5.3	2434	22	AAU34339	Staphylococcus aure
27	129	5.3	6281	22	AAU37403	Staphylococcus aure
28	128.5	5.3	1026	15	AAAR6993	real S-lyase prote
29	128.5	5.3	1026	17	AAK94014	Calobacter S-layer
30	128.5	5.3	3257	22	ABBE7502	Drosophila melanog
31	127.5	5.3	596	20	AAAY2070	Mycobacterium tuberc
32	127.5	5.3	596	23	AAE17574	Mycobacterium spec
33	127.5	5.3	599	23	AAU45599	Antigenic fusion p
34	127.5	5.3	729	22	AAO32142	Ra12-Hg-32a fusion
35	127.5	5.3	729	23	AAE17573	Mycobacterium spec
36	127.5	5.3	2541	21	AAAB1067	Human ORF ORF851
37	127	5.3	676	21	AAAS9271	Human huntingtin-l
38	127	5.3	2123	22	AAE0701	Moraxella catarrhina
39	126.5	5.2	669	23	AAAY00141	Enterococcus faeca
40	126.5	5.2	669	20	ABPA3360	E faecalis EF069 a
41	126.5	5.2	1026	18	AAAG7490	Calobacter crescent
42	126.5	5.2	1026	21	AAAY4757	Calobacter crescent
43	126.5	5.2	1638	20	AAAY00138	Enterococcus faeca
44	126.5	5.2	1638	20	AAAY00140	Enterococcus faeca
45	126.5	5.2	1638	20	AAAY00142	Enterococcus faeca

ALIGNMENTS

XX	RESULT 1
XX	AAE12573
ID	AAE12573 standard; Protein; 486 AA.
XX	
AC	AAE12573;
XX	
DT	03-JAN-2002 (first entry)
XX	
DE	Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 encoded protein.
XX	
KM	Conserved Effector Loc1, CEL; cytostatic; antibacterial; gene therapy;
KM	Exchangeable Effector Loc1, EE1; disease resistance; transgenic plant;
XX	eukaryotic cell death; cancer.
XX	
OS	Pseudomonas syringae.
XX	
PN	MO200175066-AA. ←
XX	
PD	11-OCT-2001.
XX	
PF	03-APR-2001; 2001WO-US10698.
XX	
PR	03-APR-2000; 2000US-194160P.
PR	11-AUG-2000; 2000US-224604P.
PR	17-NOV-2000; 2000US-249548P.
XX	
PA	(CORR) CORNELL RES POUND INC.
PA	(UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
PA	(UYNE-) UNIV NEBRASKA.
XX	
PI	Collmer A, Alfano JR, Charkowski AO;
XX	
DR	WPI; 2001-639361/73.

Scanned in. - NO 1022

NO: US (2001-73) NO 1022

Not worth part
of diff. in units

QY 181 LSASPDRLHDMAASVSRHSPLARQVLDTCVAVQVTSARNNAVETVLAPALASRPVQGA 240
 DB 181 LSASPDRLHDMAASVSRHSPLARQVLDTCVAVQVTSARNNAVETVLAPALASRPVQGA 240
 QY 241 VDLGVSMAGGLAANAGGNRLISVQSRDHORGALVTLGDKXKPKAQLSENDMLEAYKA 300
 DB 241 VDLGVSMAGGLAANAGGNRLISVQSRDHORGALVTLGDKXKPKAQLSENDMLEAYKA 300
 QY 301 IKSASYSGALNAGKRMAGPLDVAITDAMGAVRSLSVSSLTONGGLAGFAGVGLQK 360
 DB 301 IKSASYSGALNAGKRMAGPLDVAITDAMGAVRSLSVSSLTONGGLAGFAGVGLQK 360
 QY 361 MATKNITDPATKAAVSQLTNLGSAVAFAGWTTAALTTPDAVKKAESEFIQDTVKSTAST 420
 DB 361 MATKNITDPATKAAVSQLTNLGSAVAFAGWTTAALTTPDAVKKAESEFIQDTVKSTAST 420
 QY 421 TGYVADQTVLAKTVKDMGSEALITHTGASLRNTVNNLRORPARADIEEGSTAASPEIR 480
 DB 421 TGYVADQTVLAKTVKDMGSEALITHTGASLRNTVNNLRORPARADIEEGSTAASPEIR 480
 QY 481 FRPMRS 486
 DB 481 FRPMRS 486

RESULT 3

AAE12603
 ID AAE12603 standard; Protein; 487 AA.

XX AAE12603;
 AC
 XX

DT 03-JAN-2002 (first entry)
 XX

DE Pseudomonas syringae pv. tomato strain DC3000 HopPtoA2 protein.

XX Conserved Effector Loci; CEL; cytosolic; antibacterial; gene therapy;
 KM Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
 KW eukaryotic cell death; cancer.

XX Pseudomonas syringae.
 OS

XX WO200175066-A2.
 PN

XX 11-OCT-2001.
 PD

XX 03-APR-2001; 2001MO-US10698.
 PF

XX 03-APR-2000; 2000US-194160P.
 PR

XX 11-AUG-2000; 2000US-224604P.
 PR

XX 17-NOV-2000; 2000US-249548P.
 PR

XX (CORR) CORNELL RES FOUND INC.
 PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.

PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
 PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.

XX Collmer A, Alfano JR, Charkowski AO;
 PI

XX WPI; 2001-639361/73.
 DR

XX N-PSDB; AAD20438.
 DR

XX New nucleic acid molecules encoding proteins or polypeptides of
 PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
 PT genomic sequences, for imparting disease resistance to plants

XX Claim 8, Page 65-66; 217p; English.
 PS

XX The invention relates to an isolated nucleic acid molecule comprising a
 CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
 CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
 CC genomic sequences. CEL and EEL DNA are useful for imparting disease
 CC resistance to a plant, by transforming a plant cell with the nucleic acid
 CC and regenerating a transgenic plant from the transformed plant cell;

CC where the transgenic plant expresses a heterologous DNA molecule under
 CC conditions effective to impart disease resistance, or by treating a plant
 CC with an isolated protein or polypeptide, by applying the protein or
 CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
 CC which secretes the protein or polypeptide, under conditions effective to
 CC impart disease resistance to the treated plant. CEL and EEL proteins
 CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
 CC Pseudomonas protein into a eukaryotic cell under conditions effective to
 CC cause cell death. CEL and EEL proteins are also useful for treating a
 CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
 CC cancer cells of a patient under conditions effective to cause death of
 CC cancer cells, and thus treating the cancerous condition. The method
 CC further involves administering a targeted DNA delivery system
 CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
 CC to the patient, where the targeted DNA delivery system delivers the
 CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
 CC is expressed in the cancer cells. The present sequence is
 CC Pseudomonas syringae pv. syringae HopPtoA2 protein.
 XX

SQ Sequence 487 AA;

Query Match 77.5%; Score 1869.5; DB 22; Length 487;

Best Local Similarity 78.4%; Pred. No. 1.4e-147; Indels 1; Gaps 1;

Matches 381; Conservative 33; Mismatches 71; Indels 1; Gaps 1;

QY 1 MHINRVOQPPVATDSFRTASDASLSVSRVSSPDQREINADYLTIDHVPAAHKL 60

DB 1 MHINQSAQPPGVAMSEFRTRSDASLSVSRVSTTSCDQALTDYLGHNPAARFS 60

QY 61 PADSADGOAADVHNAQITALLITRASRLHFEGEPTATIDTPAKAKDLRLATTSGAL 120

DB 61 VIGSPDERDPAALAHNEQIDALVETPRANRLYSEGETPATIETPAKKEKPRLLATTSSAF 120

QY 121 RAPPAMASLIQYMPAINKGDWLPAPLKEPLTSLGASGANDQVGTKMDADATGDLHY 180

DB 121 ENTTPFAASVLYQMPAINKGDWLPAPLKEPLTSLGASGANDQVGTKMDADATGDLHY 180

QY 181 LSASPDRLHDMAASVSRHSPLARQVLDTCVAVQVTSARNNAVETVLAPALASRPVQGA 240

DB 181 LSTSPDKLHDMAASVSRHSPLARQVLDTCVAVQVTSARNNAVETVLAPALASRPVQGA 240

QY 241 VDLGVSMAGGLAANAGGNRLISVQSRDHORGALVTLGDKXKPKAQLSENDMLEAYKA 300

DB 241 VDFGVSTAGGLVANAGGDMLSVQSRDQLRGGAFLVGMKDKPKAALSBETDMLDAYKA 300

QY 301 IKSASYSGALNAGKRMAGPLDVAITDAMGAVRSLSVSSLTONGGLAGFAGVGLQK 360

DB 301 IKSASYSGALNAGKRMAGPLDVAITDAMGAVRSLSVSSLTONGGLAGFAGVGLQK 360

QY 361 MATKNITDPATKAAVSQLTNLGSAVAFAGWTTAALTTPDAVKKAESEFIQDTVKSTAST 420

DB 361 MATKNITDPATKAAVSQLTNLGSAVAFAGWTTAALTTPDAVKKAESEFIQDTVKSTAST 420

QY 421 TGYVADQTVLAKTVKDMGSEALITHTGASLRNTVNNLRORPARADIEEGSTAASPEIR 480

DB 421 TGYVADQTVLAKTVKDMGSEALITHTGASLRNTVNNLRORPARADIEEGSTAASPEIR 480

QY 480 FRPMRS 486

DB 480 FRPMRS 486

XX Collmer A, Alfano JR, Charkowski AO;
 PI

XX WPI; 2001-639361/73.
 DR

XX N-PSDB; AAD20438.
 DR

XX New nucleic acid molecules encoding proteins or polypeptides of
 PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
 PT genomic sequences, for imparting disease resistance to plants

XX Claim 8, Page 65-66; 217p; English.
 PS

XX The invention relates to an isolated nucleic acid molecule comprising a
 CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
 CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
 CC genomic sequences. CEL and EEL DNA are useful for imparting disease
 CC resistance to a plant, by transforming a plant cell with the nucleic acid
 CC and regenerating a transgenic plant from the transformed plant cell;

CC Enzyme; flavour; aroma; texture; nutritional; dairy manufacture; therapy;

KW fermentation process; anti-infection; rotavirus infection; heart disease;
KW infantile diarrhoea; lactose digestion; anti-cancer; autoimmune disorder;
KW anti-mutagenesis; immune system modulation; allergy; Helicobacter pylori;
KW antihypertensive effect; urogenital infection; hepatic encephalopathy;
KW bowel syndrome; endocarditis; transgenic microbe; outer membrane protein;
KW rompA.

XX Lactobacillus rhamnosus HN001.

OS WO200212506-A1.

XX 14-FEB-2002.

XX 08-AUG-2001; 2001WO-NZ00160.

XX 08-AUG-2000; 2000US-0634238.

XX 28-NOV-2000; 2000US-0724623.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX (VIAL-) VIALACTA BIOSCIENCE NZ LTD.

XX Glenn M, Hayukkala IJ, Blokeberg LN, Lubbers MW, Dekker J;

XX Christensson AC, Holland R, O'Coole PW, Reid JR, Coolbear T;

XX WPI; 2002-241760/29.

XX N-PSDB; AAD31881.

XX New polynucleotides and polypeptides from Lactobacillus rhamnosus,
XX useful in e.g. improving the flavor, aroma, texture and health-related
XX benefits of milk-derived products, or in increasing properties of
XX microbes -

XX Claim 2; Fig 64; 257pp; English.

XX The present invention relates to a new isolated polynucleotide comprising
XX a sequence present in Lactobacillus rhamnosus strain HN001 and encoding a
XX polypeptide capable of modifying the flavour, aroma, texture, nutritional
XX and health benefits of milk-derived products, and/or survivability of
XX microbes in dairy manufacturing processes. The polynucleotides are useful
XX for improving the properties of microbes used in the manufacture of milk-
XX derived products such as cheeses, yogurt, fermented milk products, sour
XX milks and buttermilk; in modifying the flavour, aroma, texture and health
XX related benefits of milk-derived products and in increasing the survival
XX of microbes during industrial fermentation processes. The bacteria may be
XX used to increase resistance to enteric pathogens and anti-infection
XX activity, including treatment of rotavirus infection and infantile
XX diarrhoea; aid in lactose digestion; as anti-cancer and anti-mutagenesis;
XX liver cancer reduction; reduction of small bowel bacterial overgrowth;
XX immune system modulation and treatment of autoimmune disorders and
XX allergies; treatment of allergic responses to foods; reduction of blood
XX lipids and prevention of heart disease; antihypertensive effect;
XX prevention and treatment of urogenital infections, Helicobacter pylori,
XX or hepatic encephalopathy; treatment of inflammatory bowel disorder and
XX irritable bowel syndrome; modulation of endocarditis; and for improved
XX protein and carbohydrate utilization and conversion. The transgenic
XX microbial population can be administered to a mammal as an anti-
XX carcinogenic agent. The present sequence is Lactobacillus rhamnosus
XX outer membrane protein rompA.

XX Sequence 1463 AA;

XX Query Match 6.1%; Score 146; DB 23; Length 1463;

XX Best Local Similarity 22.5%; Pred. No. 0.0056;

XX Matches 112; Conservative 72; Mismatches 219; Indels 94; Gaps 20;

Qy 14 ATDSFRTASDASLASSVR---SVSSD--QOREINAIADYL-----TDHVFAAKLPPA 62

Db 996 ASEAAKASNSAATAAAGFSAASDSEQAKTAASADVASSAASTANASAAASAT 1055

Qy 63 DSADGQAADVHNAQITAIETR-ASRLHFEGETPATIADTFAK-----AEKLDPLATTT 116

Db 1056 KAGDSKAAAGFSSAAASAAASAGKAEAVASAAASAAASDSDVASSAASAAAGFDKAASAA 1115

Qy 117 SGALRATPEAMASILQYMQPAINKGDMWLPAPLKLPLISGALSGAMDQYTKMMDRATG 176
Db 1116 EGAASSAASAAAS-----SAAACG-----TRGGASSSASEAG-----QASTA 1152
Qy 177 DLHYLSASDRLHDMAASVKRHSPLARQVLDTGVAVQVYSARNNAVRTVLAPALASRPA 236
Db 1153 TSVYASAASSASE--AGSYAHQAGSSASE--TGHASSATSQASA-----ASSAASRYP 1203
Qy 237 VQGAVIDGVSMAGLANAGFGRNLLISVQRDHRQGGALVGLKDKPKAQLSENDWLE 296
Db 1204 SDSGIQSDVSIASSAASATAS-----SAAASAAQSEASTASSAASHASEQASIASSEDDVVS 1257
Qy 297 AYKAIKASYSYGALNAGK-----RMAGLPIDMATDANGAVRSLSVSASSLQTQGLALAGGF 352
Db 1258 SSAASVASSAASAAASAAKAGNSAAGIYSHAASAAASAAKSAESQASSAASAAAA----- 1312
Qy 353 AGVGKLEQEMATKNITDPATKAASVQLTNLAGSAAVAFAGWTTAALTTPAV-----KKA 405
Db 1313 -----SDDSVASSAASAAALSDDAKASSAADVASSATTAAISSATSADQSATGSTA 1363
Qy 406 ESFIQDTVKSTASTTGYV-ADOT-----VKLAKTVKDMGGEAITHTGASLRNTVNN 456
Db 1364 GSHI---LPSTGGETTGISPGQTPTQTQTKPTQTKPTQAGOT-TQTG-SLPQTDHA 1418
Qy 457 LRQEPAREADIEGGTA 473
Db 1419 GRHMLPOTGDDASGTS 1435

RESULT 5

AAR27745

ID AAR27745 standard; Protein; 1822 AA.

XX AC AAR27745;

XX DT 04-MAR-1993 (first entry)

XX DE Extracellular factor related protein.

XX KW EP*; detection; prevention; screening; diagnostic.

XX OS Streptococcus suis type II (non-pathogenic).

XX FH Key Location/Qualifiers

FT Peptide 1..46

FT Peptide /note= "signal peptide"

FT Peptide 47..1822

FT Region 858..861

FT Region /note= "mature peptide"

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 934..937

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 999..1002

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1075..1078

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1264..1267

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1362..1365

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1438..1441

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1514..1517

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1590..1593

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1666..1669

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

FT Region 1741..1744

FT Region /note= "repetitive Asn-Pro-Asn-Leu sequence"

XX PN WO9216630-A.

PD 01-OCT-1992.
 XX 19-MAR-1992; 92WO-NL00054.
 XX 21-MAR-1991; 91NL-0000510.
 PR (DIER-) CENT DIERGENESKUNDIG INST.
 PA Smith HB, Vecht U;
 PI WPI; 1992-349215/42.
 DR N-PSDB; AAQ29471.
 XX Deoxyribonucleic acid encoding virulence characteristic of
 PT Streptococcus suis - useful for antibody and polypeptide for
 PT diagnosing and preventing infections in pigs and humans
 XX
 PS Claim 9; Fig 1b; 86pp; English.
 CC The sequence is that of the extracellular factor related protein
 CC from Streptococcus suis type II (non-pathogenic) which allows the
 CC detection and the prevention of infections by S. suis in a more
 CC effective manner than was previously possible. It facilitates
 CC screening of e.g. pigs and elimination of infected and carrier pigs
 CC can then be carried out. The new diagnostic tests can distinguish
 CC between avirulent and virulent strains. It may be used in the prodn.
 CC of a vaccine. See also AAR27744 and AAR27746.
 CC
 SQ Sequence 1822 AA;

Query Match 6.0%; Score 144.5; DB 13; Length 1822;
 Best Local Similarity 21.2%; Pred. No. 0.01;
 Matches 109; Conservative 81; Mismatches 208; Indels 115; Gaps 22;

QY 14 ATSFRTASDASLASVSVSSQOREINAD-----YLTG 51
 DB 1222 ARPAVELAKKEKKEKRTREEBETATVEKLAEDTRKAIEDNENLSEDEKQAEIKKLTG 1281
 QY 52 HVEFAHKLPLPADSAD-----GQAADVHNAQITALLIETRAS--RLHF--EGETPAT 98
 DB 1282 AV-AKTLATRDADNADKRTQAEKRAQALADLEKAKETOKIADKAIIDLITLIVKDELEAT 1340
 QY 99 IADTFAKAEKLDRLATTTGALRATPFAASLLQWOPAINKGDMPLPAKPLTPLISGA 158
 DB 1341 KQD--AKT-KIAKDAAKAEKAIASNP-----NLTD--AEKKTFTDAVDAB 1380
 QY 159 LSGAMPQVGRKMDRATGDLHYLSASPDRLHDMAASVKHSPSLAQVLDTGAVQTYS 218
 DB 1381 VAKANDAIISA-----ATSS-----PADVQKEEDAGVAIAEDVLDAAKQDAKKNKIAKDAAA 1430
 QY 219 ARNAVRTVLALPALASRPVAVGAVDLGVSMAGLAANAGFGNRLLSVQSRDHORGAL--- 275
 DB 1431 AKKALIGSNPLTDEKKTFTDVADEYAKANDAIISA--TSPADVQKEEDAGVAIAED 1487
 QY 276 VLGLKKEPPAOLSEENDMLEAYKAIKSA-----SVSGAALNMGKRMAGPLDMATDMAG 330
 DB 1488 VLDAKQDAKKNKIAKESD-----AAKSAIDANPNLDAKESAKKAVDADAKAATDAID 1541
 QY 331 AVNSLVSASSLTONGLALAGFAGVGLQEMATNTITDPATKAASVOLTMLAGSAAVFAG 390
 DB 1542 ASTSPEDAQAEDEK-----GVGSI-----ADVDLDAKQDAKKNKIAKEVAAA----- 1583
 QY 391 WTTAALTTDPAVKKAE-SFIQDTYKSPASTTGGVADQYKL-AKTYKMGGEALITHTGA 448
 DB 1584 --KEALIDANPNLSDAEAKKAVDADAKAATDAIDASTSPEDAQAEDEK-----VGS 1635
 QY 449 SLRTVNNLQ-----RPAEADIEEGTAAPS 477
 DB 1636 IADVDLDAKQDAKKNKIAKESDAKSAIDANPN 1668

RESULT 6
 AAY32068

ID AAY32068 standard; Protein; 600 AA.
 XX AC AAY32068;
 XX 17-JAN-2000 (first entry)
 XX Mycobacterium tuberculosis antigen fusion protein Mtb61F.
 DE Mycobacterium tuberculosis antigen fusion protein Mtb61F.
 XX Tuberculosis; antigen; fusion protein; Mtb61F; TBH9; DPV; MTI;
 KM diagnosis; therapy; vaccine; immunogen.
 XX Mycobacterium tuberculosis.
 OS
 XX WO951748-A2.
 PN
 PD 14-OCT-1999.
 XX
 XX 07-APR-1999; 99WO-US07717.
 XX
 XX 07-APR-1998; 98US-0056556.
 PR 30-DEC-1998; 98US-0223040.
 XX
 PA (CORI-) CORIXA CORP.
 PI Skelky YAM, Alderson M, Campos-Neto A;
 DR WPI; 1999-601610/51.
 DR N-PSDB; AA220203.
 PT New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis -
 XX
 PS Claim 1; Fig 10A-B; 83pp; English.

CC This sequence represents a recombinant Mycobacterium tuberculosis
 CC tri-antigen fusion protein, termed Mtb61F, composed of the antigens
 CC TBH9, DPV and MTI. The fusion protein is expressed in host cells
 CC using a vector carrying a polynucleotide (see AA220203) comprising
 CC the coding sequences for the 3 antigens. The invention provides
 CC fusion proteins (see AAY32059-71) containing at least 2 M.
 CC tuberculosis antigens. The new fusion proteins and polynucleotides
 CC encoding them are useful as vaccines for preventing tuberculosis
 CC (claimed), for diagnosis (via in vitro assays or intradermal skin
 CC tests for detection of anti-M. tuberculosis antibodies), monitoring
 CC of disease progression, and treatment of tuberculosis. They are
 CC more effective immunogens than mixtures of the individual protein
 CC components.
 CC
 SQ Sequence 600 AA;

Query Match 5.9%; Score 143; DB 20; Length 600;
 Best Local Similarity 23.0%; Pred. No. 0.0023;
 Matches 117; Conservative 58; Mismatches 165; Indels 168; Gaps 25;

QY 20 TASDASIASSVSVSSDQOREINADYLTDFVFAHKLPLPADSADGQA----- 69
 DB 88 TAGAEITTAQYAV-----AAAAYETAGLT---VPPVIAENRBEMLILATNL 134
 QY 70 -----AVDVHNAQITALLIETRASRLHFEGETPATIADTFPAKAEKLDRLATTTGAL--- 120
 DB 135 LGONTALIAVNEAEYEMMWDAAAFGYAAATATATATLLPREBAPEM--TSAGGLLEQ 192
 QY 121 -----RATPPAAS-LLQYKQPIKNGDMPLPAKLELPLISALSGAMQVQ-----T 168
 DB 193 AAAVEESDTPAANNQMNPNVPAALQO--LAQPTQGTTP--SSKLGGLKWTVPSPHSPTIS 247
 QY 169 KMDRATGDLHYLSASPPDRLDMAASVKHSPSLARQVLDTGAVQVTSARAAVFTVLA 228
 DB 248 NWSMANNHNS-WTNSGVSTNTLSSMLKGFAPAAALQ-----AVQT-AAQNGVR----- 235
 QY 229 PALASRPVAVGAVDLGVSMAGLAANAGFGNRLLSVQSRDHORGALVLGLKKEPPAOL 288
 DB 296 -AMSTLGSSTLGSGLG-----GVVAANLIG--RAASVGS-----L 326

QY 289 SEENDWLEAYKAISAS-----YSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASS 340
 Db 327 SVPQAWAANAQAVTPAARALPLTSLTSAERGPQGLGLFVG----- 369
 QY 341 LTONGALAGGFAGV-----GKL-----QEMATKNITDP- 369
 Db 370 --QMGARAGGSLGVLVPRPPYVMPHSPAGKLDPDVDAVINTTCNYGQVVAALNATDPG 427
 QY 370 -----ATKAASVQLNLNAGSAAVFAGWTTAALTTPPAVKKAESFIQDTVKSTASSTGY 423
 Db 428 AAQFNASPVQAQSYLRNFLAAPPQRAAAQQL---QAVPGAQYI-GLVESVAGSCNNY 483
 QY 424 VADQTVKLAKTVKMGGEAITHTGASLR 451
 Db 484 E-----LMTINYQFGVDVAH-GAMIR 503
 RESULT 7
 AAU74597 standard; Protein; 600 AA.
 AC AAU74597;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Antigenic fusion protein TbH9-DPV-MTI (Mtb61f).
 XX
 KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;
 KW tuberculostatic; immunogen; vaccine; TBH9-DPV-MTI; Mtb61f.
 XX
 XX Chimeric - Mycobacterium tuberculosis.
 OS
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 301
 FT /note= "Encoded by GC"
 XX
 XX US2002009459-A1.
 XX
 XX 24-JAN-2002.
 XX
 XX 07-APR-1999; 99US-0287849.
 XX
 XX 13-MAR-1997; 97US-0818112.
 XX 01-OCT-1997; 97US-0942578.
 XX 18-FEB-1998; 98US-0025197.
 XX 07-APR-1998; 98US-0056556.
 XX 30-DEC-1998; 98US-0223040.
 XX
 XX (REED/) REED S G.
 XX (SKEI/) SKEIKY Y A.
 XX (DILL/) DILLON D C.
 XX (ALDE/) ALDERSON M.
 XX (CAMP/) CAMPOS-NETO A.
 XX
 XX Read SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
 XX
 XX WPI; 2002-171134/22.
 XX N-PSDB; ABK14137.
 XX
 XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
 PT diagnosing, treating or preventing M. tuberculosis infection,
 PT particularly as vaccine for treating or preventing tuberculosis -
 XX
 XX Claim 1; Fig 10; 62pp; English.
 XX
 XX The invention relates to a purified polypeptide which induces an immune
 CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
 CC useful for diagnosing, treating or preventing M. tuberculosis infection,
 CC particularly tuberculosis infection. In particular, the polypeptides are
 CC useful as a vaccine formulation with an adjuvant to afford long-term
 CC protection in animals against the development of tuberculosis. The
 CC protein coding sequence may be used to encode a protein product for use

CC as an immunogen to induce and/or enhance an immune response to M.
 CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
 CC of the invention.
 XX
 SQ Sequence 600 AA;
 Query Match 5.9%; Score 143; DB 23; Length 600;
 Best Local Similarity 23.0%; Pred. No. 0.0029;
 Matches 117; Conservative 58; Mismatches 165; Indels 168; Gaps 25;
 QY 20 TASDASLASVSRSVSDQOREINAIADYLDHVFAAHKLPPADSDQQA----- 69
 Db 88 TAGQAEITAAQVRV-----AAAYETAYGLT---VPPPVIAENRLEMLLIATNL 134
 QY 70 -----ADVHNAQITALITETASRLHFEGETPATIADTTFAKAEKLDRLATTTSGAL--- 120
 Db 135 LGQNTPAIAVNEAEYGEWMAQDAAMFEGYAAATATATATLLPFEAPEM--TSAGGLEQ 192
 QY 121 -----RATPEAMAS-LLOYMQPAINKDMLPAPLKPLTPLISGALSAMDQVG-----T 168
 Db 193 AAAVEEASDTAAANQLMNNVPQALQQ---LAQPTQGTTP--SSKLGGLWKTVPSPSPIS 247
 QY 169 KMDRATGDLHYLSASPDRLHDAMAASVKRHSPLARQVLDTGAVQVYSGARNAVTVLA 228
 Db 248 NMYSMANNHMS-WTNSGVSMNTLSSMLKGFAPAAAQ-----AVQT-AAQNGVR----- 295
 QY 229 PALASRPVQGVAVDLGVSMAGGLAANAGFGRNLLSVOSRDHQRGGALVLGLKDKPEKAQL 288
 Db 296 -AMSSLGSSLSGSSGLG---GGVAANLG---RAASVGS-----L 326
 QY 289 SEENDWLEAYKAISAS-----YSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASS 340
 Db 327 SVPQAWAANAQAVTPAARALPLTSLTSAERGPQGLGLFVG----- 369
 QY 341 LTONGALAGGFAGV-----GKL-----QEMATKNITDP- 369
 Db 370 --QMGARAGGSLGVLVPRPPYVMPHSPAGKLDPDVDAVINTTCNYGQVVAALNATDPG 427
 QY 370 -----ATKAASVQLNLNAGSAAVFAGWTTAALTTPPAVKKAESFIQDTVKSTASSTGY 423
 Db 428 AAQFNASPVQAQSYLRNFLAAPPQRAAAQQL---QAVPGAQYI-GLVESVAGSCNNY 483
 QY 424 VADQTVKLAKTVKMGGEAITHTGASLR 451
 Db 484 E-----LMTINYQFGVDVAH-GAMIR 503
 RESULT 8
 AAM78810
 ID AAM78810 standard; Protein; 1134 AA.
 XX
 AC AAM78810;
 XX
 XX 06-NOV-2001 (first entry)
 XX
 XX Human protein SEQ ID NO 1472.
 DE
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 XX WO200157190-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 05-FEB-2001; 2001WO-US04098.
 XX
 XX 03-FEB-2000; 2000US-0496914.
 XX 27-APR-2000; 2000US-0560875.
 XX 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
 PI Zhao QA, Wang Z, Wehrman T, Ren F, Chen R, Wang ZW,
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
 DR MPI; 2001-476283/51.
 DR N-PSDB; AAKS1943.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 20, Page 3747-3749; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK60302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM60020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX

SO Sequence 1134 AA;

Query Match 5.9%; Score 142.5; DB 22; Length 1134;
 Best Local Similarity 23.4%; Pred. No. 0.0077;
 Matches 122; Conservative 70; Mismatches 248; Indels 81; Gaps 22;

QY 19 RTSDASGLASSSVRSVSDQREINADYLTTHVFAHKL--PPADSAGQAAVDVHNA 76
 DB 267 RDIEQSLAAVVS--QSLATRDIDISVEALQEQITVVOBIGHIDPIATNAAGGAQLGH-- 323
 QY 77 QITALETRASRLHFEGETPATIADTFAXKAEKDRLATTTSGALRATPPAMASLLQYMP 136
 DB 324 KVTQLAS-----YFE--PLILAVGVASKILDHQQOMT--VIDQTKTAAESALQ--MLY 371
 QY 137 AINKGDWLPAPLPKPLTPPLISGA--LSGAMPDQVTKMDRATGDLHYLSASPPRLHDMAA 194
 DB 372 AAEKGGNPFKAQHTHDAITBAQOLMEKAVDDI--MTVLNEASBVGVLGVGWDIAEAMSK 430
 QY 195 SVKRHSLSLRQVLDGTGVAVQTSARNAV-----RTVLAALASRPVQAVDGLG--VS 246
 DB 431 LDGETPEPKGTFTDXYQTTVTKYSKSIATVAQEMWTKSVNPEELIGLASQMTSDYGHLA 490
 QY 247 MAGGLAA-----NAGGNLLSVQSRDH-----QKGGALVLGKXKERAQLSE--END 293
 DB 491 FQGGMAATAPPEIRGQIR--TRVQDIGHGCIPLVQAGALQVCPSTSYTRRELIECARA 549
 QY 294 WLEAYKAIKSASVSG-----AALNAGKRMAGLPMDMATDMGAVRSIVSASL-----T 342
 DB 550 VTEKVSIVLAFALQAGNKGCTQACTAATAVAGGIADLDTTTFMFAAGTLNENSTFADHR 609
 QY 343 QNGSLAAGFAGVCKLOEMATKNTTDPATQAASQTLNLAGSAAPVFGWTTAALTDP-- 400
 DB 610 ENLTKAKALVEDTKLIVSGAASPPDLAQAASSAATITQLAEVVVLGAASLSDDPET 669
 QY 401 -----AVKTAESPIDQTVSTASTTGVADQTV-----KLAKTVDMG 439
 DB 670 QVVLINAIKQVAKALSLDISITKGAASKPVDDPSMYQLKGAAYMTVNTSLTKTAAVE 729

QY 440 GEAITHTGASLRTVNNLRQRP--READIEEGCTAASPS 478
 DB 730 DEATRGTRRA--LEATIECIRKQELTVFSKQVPE--KTSPEE 767

RESULT 9

AAM39293
 ID AAM39293 standard; Protein; 1144 AA.

AC AAM39293;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 2438.

XX Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

DR MPI; 2001-442253/47.
 DR N-PSDB; AAI58449.

PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -

PS Example 4; SEQ ID NO 2438; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM36642-AAK42213) with nocotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression.
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

SO Sequence 1144 AA;

Query Match 5.9%; Score 142.5; DB 22; Length 1144;

Best Local Similarity 23.4%; Pred. No. 0.0078;		Matches 122; Conservative 70; Mismatches 248; Indels 81; Gaps 22;	
Qy	19	RTASDASLASSSVRSVSSDOOREINAIADYLTDFHFAAHKL--PPADSADGQAADVHNA 76	
Db	277	RDIEQASLAASV--QSLATRDIDISVEALQEQLTSSVQVEIGHLIDPIATAARGEAAQLGH-- 333	
Qy	77	QITALITRASRLHFEGETPATIADTFKAELKDLRLATTTSGALRATPPFAMASLLQYNQP 136	
Db	334	KVTQLAS-----YFE---PILAAVGVASKILDHQOQMT--VLDQTKTLAESALQ-MLY 381	
Qy	137	AINGKDWLPAPLPLTPLISGA--LSGAMDQVGTQMDRATGDLHYLSASPDRLHDMAA 194	
Db	382	AAKEGGGNPKAQHTHDAITEAAQMKAEVDDI-MVTLINEAASEVLGVGMVDAIAEANSK 440	
Qy	195	SVKRHSPLARQVLDTGVAQVQTSARNV-----RTVLAPALASRPVQGVADLG-VS 246	
Db	441	LDEGTPPEPKGTFVDYQTTVVKYSKSIATVQEMMTKSVTNPEELGGLASQMTSDYGHLA 500	
Qy	247	MAGGLAA-----NAGFNRLLSVQSRDH-----QRGALVLGLKDKPEKPAQLSE-END 293	
Db	501	FOGQMAAATAEPPEIGFQIR-TRVQDLGHGCTFLVQKAGALQVCPTDSYTKRELIECARA 559	
Qy	294	WLEAYKAIKSASYSG-----AALNAGKEMAGLPDMDATDAMGAVRSLSVSSSL-----T 342	
Db	560	VTEKVSLSLALQAGNKGTQACITAAATAVSGIITADLDTTINFATAGTINAENSETFADHR 619	
Qy	343	QNGLALAGGFAGVKGLOEMATKNIITDPATKAASVQLTNLAGSAVFAAGTAAALTDTP-- 400	
Db	620	ENILKTAALVEDTKLIVSGAASSTPKLAQAQSSAAITITQLAEVVKLGAASLGSDDPET 679	
Qy	401	-----AVKKAESFIQDTVKVSTASTTGYVADQTV-----KLAKTVKDMG 439	
Db	680	QVVLINAIKDVAKALSDLISATKGAASKPVDPSMYQLKGAAKVMVTNVTSLTKTRAVE 739	
Qy	440	GEATHTGASLRNTVNNLRORPA--READIEGGTAAASPSE 478	
Db	740	DEATRGTRA-LEATIECIKQELTVFQSKDVPE--KTSSPEE 777	
RESULT 10			
ID	AA39312	standard; Protein; 1165 AA.	
AC	AA39312;		
XX	22-OCT-2001	(first entry)	
DE	Human polypeptide	SEQ ID NO 2457.	
XX	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
XX	leukaemia.		
OS	Homo sapiens.		
XX	WO200153312-A1.		
XX	26-JUL-2001.		
XX	26-DEC-2000; 2000WO-US34263.		
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
<p>XX (HYSE-) HYSEQ INC.</p> <p>Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J; Zhao QA, Zhou P, Goodrich R, Drmanac RT;</p> <p>WPI; 2001-442253/47. N-PSDB; AAI58468.</p> <p>Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -</p> <p>Example 4; SEQ ID NO 2457; 10078pp; English.</p> <p>The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.</p> <p>Note: The sequence data for this patent did not form part of the printed specification.</p> <p>Sequence 1165 AA;</p> <p>Query Match 5.9%; Score 142.5; DB 22; Length 1165; Best Local Similarity 23.4%; Pred. No. 0.008; Matches 122; Conservative 70; Mismatches 248; Indels 81; Gaps 22;</p> <p>Qy 19 RTASDASLASSSVRSVSSDOOREINAIADYLTDFHFAAHKL--PPADSADGQAADVHNA 76 Db 298 RDIEQASLAASV--QSLATRDIDISVEALQEQLTSSVQVEIGHLIDPIATAARGEAAQLGH-- 354</p> <p>Qy 77 QITALITRASRLHFEGETPATIADTFKAELKDLRLATTTSGALRATPPFAMASLLQYNQP 136 Db 355 KVTQLAS-----YFE---PILAAVGVASKILDHQOQMT--VLDQTKTLAESALQ-MLY 402</p> <p>Qy 137 AINGKDWLPAPLPLTPLISGA--LSGAMDQVGTQMDRATGDLHYLSASPDRLHDMAA 194 Db 403 AKEGGGNPKAQHTHDAITEAAQMKAEVDDI-MVTLINEAASEVLGVGMVDAIAEANSK 461</p> <p>Qy 195 SVKRHSPLARQVLDTGVAQVQTSARNV-----RTVLAPALASRPVQGVADLG-VS 246 Db 462 LDEGTPPEPKGTFVDYQTTVVKYSKSIATVQEMMTKSVTNPEELGGLASQMTSDYGHLA 521</p> <p>Qy 247 MAGGLAA-----NAGFNRLLSVQSRDH-----QRGALVLGLKDKPEKPAQLSE-END 293 Db 522 FOGQMAAATAEPPEIGFQIR-TRVQDLGHGCTFLVQKAGALQVCPTDSYTKRELIECARA 580</p> <p>Qy 294 WLEAYKAIKSASYSG-----AALNAGKEMAGLPDMDATDAMGAVRSLSVSSSL-----T 342 Db 581 VTEKVSLSLALQAGNKGTQACITAAATAVSGIITADLDTTINFATAGTINAENSETFADHR 640</p> <p>Qy 343 QNGLALAGGFAGVKGLOEMATKNIITDPATKAASVQLTNLAGSAVFAAGTAAALTDTP-- 400 Db 641 ENILKTAALVEDTKLIVSGAASSTPKLAQAQSSAAITITQLAEVVKLGAASLGSDDPET 700</p> <p>Qy 401 -----AVKKAESFIQDTVKVSTASTTGYVADQTV-----KLAKTVKDMG 439 Db 701 QVVLINAIKDVAKALSDLISATKGAASKPVDPSMYQLKGAAKVMVTNVTSLTKTRAVE 760</p> <p>Qy 440 GEATHTGASLRNTVNNLRORPA--READIEGGTAAASPSE 478 Db 761 DEATRGTRA-LEATIECIKQELTVFQSKDVPE--KTSSPEE 798</p>			

RESULT 11

AAM79794

ID AAM79794 standard; Protein: 1177 AA.

AC AAM79794;

DT 06-NOV-2001 (first entry)

DE Human protein SEQ ID NO 3440.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 tissue growth factor; immunomodulatory; cancer; leukaemia;
 nervous system disorder; arthritis; inflammation.

OS Homo sapiens.

PN WO200157190-A2.

PD 09-AUG-2001.

PF 05-FEB-2001; 2001WO-US04098.

PR 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620335.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

PR 30-NOV-2000; 2000US-0728422.

DR N-PSDB; AAK52927.

DR WPI; 2001-476283/51.

PT Nucleic acids encoding polypeptides with cytokine-like activities.

PT useful in diagnosis and gene therapy -

XX Claim 20; Page 336-337; 6221PP; English.

The invention relates to polynucleotides (AAK51456-AAK51435) and the encoded polypeptides (AAM78323-AAK60302) that exhibit activity relating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAM60020) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

SQ Sequence 1177 AA;

Query Match 5.8%; Score 139.5; DB 22; Length 1177;

Best Local Similarity 23.0%; Pred. No. 0.014;

Matches 120; Conservative 69; Mismatches 250; Indels 83; Gaps 21;

QY 19 RTASDASLASSSVSSDQREINATADYLTDFVFAAHL--PPASAGQAAVDVH-N 75

DB 310 KDIEQASLAASV--QSLATRDIDIVEALQEQULTSVQVGEIGHLIDPIATPARGEAALQGHKG 368

QY 76 AQTALIEPRASRLHFEGETPATIADTFAPAKELDRILATTTSGALRATPPAMSLQYMQ 135

DB 369 TQLASYE-----PLILAAVGASKILDHQQQMT--VLDDTKTLASALQ-ML 413

QY 136 PAINKGMLPAPLKPLTPLISGA--LSGANDQGTGKMDRATGDLHYLSAPDLHDAMA 193

DB 414 YAAKEGGGNPFAQHTHDAITEAQLMKEAVDDI-MVTLNEAASEVGLVGVAVDAIAEAMS 472

QY 194 ASVGRHSPLARQVLDGVAVQVTSARNAV-----RTVLAPALASRPANQAVDVG-V 245

DB 473 KLDEGTPPEPKGTFTVDQTTVVKTSKAIATVADEMNTKSVNPEELGQLASQMTSDYGL 532

QY 246 SMAGSLA-----NAGFGRNLISVQSRDH-----ORGALVYLKDKPEKQISE-EN 292

DB 533 AFQCGMAAATAAPEEIGFQR-TRVQDLGHCIFLVQKAGALQVCPDSTYTKRELICAR 591

QY 293 DWLEAVYAIKASYSG-----AALNAGKRNAGPLMDATDMAGVRSLSVANSSTI----- 341

DB 592 AVTEKVSILVLSALQAGNKGTQACTAATAFAVSGIADIDITIMEATAGTLNAENSETPADH 651

QY 342 TONGALAGGPAGVGKLEMATKNITDPATKAIVSOLTLNAGSAAVRAGWTTALTDP- 400

DB 652 RENTLKTAKALVEDTKLVGSAASTPDKLAQAASAAITTLQLEAVVKGAAASIGSDPE 711

QY 401 -----AVKKAESPIDQTVKSTASTSTGYVADQTV-----KLAKTYKDM 438

DB 712 TQVVLIAIDVAKALSDLSATGAAKRVDPDSWQLGAAKVMVTNTSLKTKAV 771

QY 439 GGEAITHTGASLRNTVNNLRPRA--READIERGGTAASPE 478

DB 772 EDEATRGTRA-LEATIECIRQELTVFQSKDVE--KTSPEE 810

RESULT 12

AAM41079

ID AAM41079 standard; Protein: 1177 AA.

AC AAM41079;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 6010.

Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 peripheral nervous system; neuropathy; central nervous system; CNS;
 Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSBQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 Zhao QA, Zhou P, Goodrich R, Drmanac RT;

DR WPI: 2001-442253/47.
 DR N-PSDB; AAI60235.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Example 2; SEQ ID NO 6010; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AA42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX Sequence 1177 AA;
 SQ Query Match 5.8%; Score 139.5; DB 22; Length 1177;
 Best Local Similarity 23.0%; Pred. No. 0.014;
 Matches 120; Conservative 69; Mismatches 250; Indels 83; Gaps 21;
 Qy 19 RTASDASLASSSVSSDQOREINAIADYLDTHVFAHKL--PPADSADQAAVDVH-N 75
 Db 310 RDIEQASLAASV--OSLATRDSDISVEALQEQLTSTVQVEIGHLIDPIATAARGEAQLGHKG 368
 Qy 76 AQITALITRASRLHFEGETPATIADTFKAKEKLDRLATTTSGALRATPFAMASLLQYMQ 135
 Db 369 TQLASYFE-----PLILAAGVASKILDHQQOMT--VLDQTKTLESALQ-ML 413
 Qy 136 PAINKGDLWPAKPLPLTLLISGA--LSGAMDQVGTQKMDRATGDLHYLSASPDRLHDAMA 193
 Db 414 YAAKEGGGNPRAQHTHDAITEAQLMKEAVDDI--NVTLINEAASEVLGVGMVDAFAEAMS 472
 Qy 194 ASVKRHSPLARQVLDTVGAVQTYGARNV-----RTVLAPALASPAVQGVADLG-V 245
 Db 473 KLDEGTPPEPKGTFFDYQTTVVVYKSAIAVTAQEMMTKSVTNPEELGGLASQMTSDYGH 532
 Qy 246 SMAGGLAA-----NAGFGNLLSVQSRDH-----QRGGALVLGLKDKPKAQLSE-EN 292
 Db 533 AFQGMMAAATAPEPEIGFOIR--TRVDLGHGCFILVQKAGALQVCPDTSYTKRELIECAR 591
 Qy 293 DWLEAYKAIKSASYSG-----AALNAGKRMAGLPDMDATDAMGAVRSLVSASSL----- 341
 Db 592 AVTEKSVLSALQAGNKGTOACITAAVSGIADLDTTIFATAGTLNAENSETFADH 651
 Qy 342 TONGALAGGFAVGKLOEMATKNTTDPATKAASQLTNLAGSAAVFAGWTTAALTDP- 400
 Db 652 RENILKAKALVEDTKLLVSGAASFPDKLAQAQSSAATITQLAEVVKLGAASLGSDDP 711
 Qy 401 -----AVKKAESFIQDTVKSTASTTGYVADQTV-----KLAQTVKDM 438
 Db 712 TOVVLINAIKOVAKALSLLISATKGAASKPVDDPSYQKGAUKVNTVNTVSLTKTVAV 771
 Qy 439 GGEAITHTGASLRNTVNNLRQFA--READIEEGTAASPS 478
 Db 772 EDEATRGTRA-LEATIEICKQLTVPQSKDVE--KTSSPEE 810
 RESULT 13
 AAM41098
 ID AAM41098 standard; Protein; 1177 AA.
 XX
 AC AAM41098;

XX DT 22-OCT-2001 (first entry)
 XX Human polypeptide SEQ ID NO 6029.
 XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX Homo sapiens.
 OS WO200153312-A1.
 PN 26-JUL-2001.
 XX 26-DEC-2000; 2000WO-US34263.
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR N-PSDB; AAI60254.
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Example 2; SEQ ID NO 6029; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AA42213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX Sequence 1177 AA;
 SQ Query Match 5.8%; Score 139.5; DB 22; Length 1177;
 Best Local Similarity 23.0%; Pred. No. 0.014;
 Matches 120; Conservative 69; Mismatches 250; Indels 83; Gaps 21;
 Qy 19 RTASDASLASSSVSSDQOREINAIADYLDTHVFAHKL--PPADSADQAAVDVH-N 75
 Db 310 RDIEQASLAASV--OSLATRDSDISVEALQEQLTSTVQVEIGHLIDPIATAARGEAQLGHKG 368
 Qy 76 AQITALITRASRLHFEGETPATIADTFKAKEKLDRLATTTSGALRATPFAMASLLQYMQ 135
 Db 369 TQLASYFE-----PLILAAGVASKILDHQQOMT--VLDQTKTLESALQ-ML 413

QY 136 PAINKGDMPLKPLTPPLISGA--ISGAMDQVGTKMMDRATGDLHYLSASPDRLHDMA 193
 DB 414 YAAEEGGGNPKAQHTHDATIEAQLMKEAVDDI-MTTLNEAASSEVLGVGNVDIAIEMMS 472
 QY 194 ASYGRHSPLARQVLDGVAVQVTSARNV-----RTVLAPLARSRAVQGVLDG-V 245
 DB 473 KLDEGTPPEPKGTFTVDQVTVVKYSKAIAYVAQEMTKSVTNPELGGLSQMTSDGHL 532
 QY 246 SMAGGLAA-----NAGFGRNLSVQSRDH-----ORGALVVLGDKPEKQOLSE-EN 292
 DB 533 AFQGGMAAAARAEPEIGFOR-TRVQDLGHGICILVQKAGLVQCPDSTYKRELICAR 591
 QY 293 DWLEAVYAIKASASYG-----AALNAGKRWAGLPDMATDMAGVRSLSVASST- 341
 DB 592 AVTEKVELSVLSALQAGKKGTOACITATAVSGIILADLTITMPTAGTLMNENSETFADH 651
 QY 342 TONGALAGGAGVGLQEMATKNTITPATKANVSLQTLNLAGSAVAFAGTTAALTDP- 400
 DB 652 RENILKTAKALVEDTKLLVSGMASTPDKLAQMASSAATITQLAEVVYLGMAASIGSDPE 711
 QY 401 -----AVKKAESFIQDVTWSTASTTGVAADQTV-----KLAKTVKDM 438
 DB 712 TQVYLIAIKDVAKALBDSLISATGASKPYDDPSMTQLKGAAYMTNTVTSLLTKVAV 771
 QY 439 GGEAITHTGASLRNTVNNLRQRP--READIEEGGTAASPE 478
 DB 772 EDEATRGTRA-LEATIECIGQELTFVQSKDVE--KTSSPE 810

RESULT 14

AA75477
 ID AAY75477 standard; Protein; 723 AA.

XX AAY75477;

DT 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 719 protein sequence SEQ ID NO:2428.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KM antibacterial; gene therapy.

XX Neisseria meningitidis.

OS Neisseria meningitidis.

XX Neisseria meningitidis.

PN MO9957280-A2.

XX MO9957280-A2.

PD 11-NOV-1999.

XX 11-NOV-1999.

PF 30-APR-1999; 99MO-US09346.

XX 30-APR-1999; 99MO-US09346.

PR 01-MAY-1998; 98US-0083758.

XX 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

XX 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0098994.

PR 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103796.

XX 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX 25-FEB-1999; 99US-0121528.

PA (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX (CHIR) CHIRON CORP.

XX (GENO-) INST GENOMIC RES.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M,

PI Peterben J, Piazza M, Rappuoli R, Ratti G, Scalato E, Scarselli M,

PI Tettein H, Venter JC;

XX WPI; 2000-062150/05.

DR N-PSDB; AA254239.

XX Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics

XX Claim 2; Page 1164; 1453pp; English.
 PS AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 723 AA;

SQ Query Match 5.6%; Score 135; DB 21; Length 723;

Best Local Similarity 22.4%; Pred. No. 0.017;

Matches 119; Conservative 76; Mismatches 224; Indels 112; Gaps 25;

QY 14 ATDSFTASDASIASVSRSVSSDQREINAIADYLTDFVFAHKLPADSDGQAADV 73
 DB 166 ATEGAQOIKD--LALIEVKNKGTHDKALDLISGMTTGILNFAQTKNEAQAAYAFALASE 223
 QY 74 HMQIITA-LIET-----RASRLHEGTPATITADTFKAKRLRLATTTGALRATP 124
 DB 224 GSGEDTAKLIKLIKDGSGMSKDIQLGLEHVLSGLDGTFFVRMVVELPFLSLAAQQAQM 283
 QY 125 FMAAS--LQVWOPALINKGDMPLPAPL-----KPLTPPLISGAL-----SGAM 163
 DB 284 NGVGGDLVLLSLQSAANKSG-SFAEATTVQNLSTKLSPTDITGRKKRANRPDPKGV 342
 QY 164 DQVGTKMDRATD--LHYLSASPDRLHDMAASVYKHSPLARQVL-DTGVAVQVTSAR 220
 DB 343 DWIGSVVQKQNGENAVQVLS---RLADAMLVKDKQYQYKRRAAAGDTAAEQAMLK 398
 QY 221 NAVRTVLAAPLARSRAVQGVLDL-----GVSMAGLANAGFGRLLSVQSRDH 269
 DB 399 GALLAQLPDLQKGLLAATDITQIREVYASLAGVTLNDGKIAKNEA-RMLSAQAQOE 457
 QY 270 QRGALVVLGDKPEPKAQLSE-----NDWLBAVYKAIKSASYGAALNAGKRWAGLPDM 324
 DB 458 Q-----QESLAMERESLGTIVDMETSPKCL-AAEYFNATL----- 492
 QY 325 ATDMAGVRSLSVASSLTONGALAGGAGVGLQEMATKNTI-TDPATKAAVSQTLNLAG 383
 DB 493 ---ALQALTTAATAAASAM--LTTAGGKGAGFLKDVGSKALMGKASAGVAAGATPAG 547
 QY 384 SAAVFAQWTTAA--LTTDAVYKKAESFIQDVTWSTASTTGVAADQTVGLATVQDMG 440
 DB 548 GKLL--SWGKSAGSGLMNNPALVYKAGLLGMLLYSES-----LGDGTLT-----KGLRG 594
 QY 441 EALTHTGASLRNTV--NNLRQRPAREADIEEGGT---AASPSEIAPPKMS 486
 DB 595 ---TKTTPKINRLKNGIRFEPAPKREQARGVPOYLAAPSOPTDKMIS 642

RESULT 15

AAE17572
 ID AAE17572 standard; Protein; 729 AA.

XX AAE17572;

DT 22-APR-2002 (first entry)

DE Mycobacterium species MTB72F fusion protein.

XX Fusion protein; antigen; serological sensitivity; immune response;

KM tuberculosis; infection; vaccine; MTB72F; Ra12-TbH9-Ra35 protein.

XX

Mycobacterium sp.
Key Location/Qualifiers
Region 8..139
/label= Ra12_protein
Region 142..532
/label= TBH9FL_protein
Region 535..729
/label= Ra35_protein

W0200198460-A2.
27-DEC-2001.
20-JUN-2001; 2001WO-US19959.
20-JUN-2000; 2000US-0597796.
01-FEB-2001; 2001US-265737P.
(CORI-) CORIXA CORP.
Skeiky Y, Reed S, Alderson M;
WPI; 2002-147798/19.
N-PSDB; AAD28342.

Composition comprising MTB39 antigen and MTB32A antigen from
Mycobacterium species, useful for eliciting immune response in a
subject -
Claim 6; Fig 5; 136pp; English.

The present invention relates to fusion proteins containing at least two Mycobacterium species antigens, nucleotides encoding them and compositions comprising such fusion proteins. The present invention particularly relates to nucleic acids encoding fusion proteins that include two or more individual M. tuberculosis antigens which increase the serological sensitivity of sera from individuals infected with tuberculosis and methods for their use in diagnosis, prevention and treatment of tuberculosis infection. Sequences of the invention are useful for eliciting an immune response in a mammal, e.g., human, immunised with BCG. They are useful in the diagnosis, treatment and prevention of Mycobacterium infection. The fusion proteins and the polynucleotides are useful as diagnostic tools in patients infected with Mycobacterium, in vitro and in vivo assays for detecting humoral antibodies or cell-mediated immunity against M. tuberculosis, for the diagnosis of an infection or monitoring of disease progression, as immunogens to generate or elicit a protective immune response in a patient and for raising anti-M. tuberculosis antibodies in a non-human animal. Sequences of the invention are also used as vaccines. MTB32A fusion proteins of the invention are useful as in vivo diagnostic agents for intradermal skin test. The present sequence is a DNA encoding Mycobacterium species MTB72F (Ra12-TbH9-Ra35) fusion protein.

Sequence 729 AA;

Query Match 5.5%; Score 133.5; DB 23; Length 729;
Best Local Similarity 21.0%; Pred. No. 0.023;
Matches 116; Conservative 70; Mismatches 178; Indels 189; Gaps 25;

QY 20 TASDASLASSVRSVSSQOQREINAIADYLDTHVFAHKLPPADSADQQA----- 69
DB 221 TAGQAEITAAQVRV-----AAAAYETAYGLT---VPPPVIAENRAELMILIAITNL 267
QY 70 -----AVDVHNAQITALJETRAHLFEGETPATIAFAKELDRLATTTSGAL--- 120
DB 268 LGONTPTAVNAEAYGENWAQDAAMFGYAATATATATLLPFEAPEM--TSAGGLLEQ 325
QY 121 -----RATPFAMAS--LLQYMQFAINKGDWLPAPLPLTPLISGALSGAMDQVG----- 168
DB 326 AAAVEEASDTAAANGLMNVQALQ---LAQPTQGTTP---SSKLGGLWKTVPSPHRSPI 380
QY 169 KMMDRATGDLHYLSASPRLLHDMAAASVKRHSPLARQVLDTGVAVQVYSARNVRTVLA 228

DB 381 NMVSMANNHMS--MTNSGVSMNTLSSWLKGFAPAAARQ-----AVQT-AAQNGVR----- 428
QY 229 PALASREPAVQAVDLGVSMAGGLAANAGFNGRLLSVQSRDHQRGGALVLGLKDKPEKAQL 288
DB 429 -AMSSLSGLSSGLG---GGVAANILG---RAASVGS-----L 459
QY 289 SEENDWLEAYKAIKSAS-----YSGAALNAGKRWAGLPLDMATDAMGAVRSLVSASS 340
DB 460 SVPOMAAANQAVTPAARALPLTSLTSAERGPQMLGGLPVG----- 502
QY 341 LTQGLALAGGFAGVCKL-----QEMATKNITDPA-----TKAAVSQ 377
DB 503 --QNGARAGGGLSGVLKVPVPPVPHSPPAAGDIAPPALSQDRFADFPALPLDPSAMVAQ 560
QY 378 L-----TNLAGSAAVFAGWTTAALTDP-----AVKKAESFIQDTVKSTA 417
DB 561 VGPQVNVNITKLGYNNAVAG---TGIVDPNGVLTNNHVIAGATDINAFSVSGGQTYG 617
QY 418 SSTTGYVADQTVKLAK-----TVKDMGGEAITHTGASLRNTVNNLRORPAREADIEE 469
DB 618 VDVVGYDRTQDVAVLQLRGAGGLPSAIGGVAVGEPVWAMGNS----- 661
QY 470 GGTAAASPSEIPER 482
DB 662 GGQGGTPRAVPGR 674

Search completed: January 2, 2003, 15:18:52
Job time : 40.4645 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:14:20 ; Search time 9.49024 Seconds
(without alignments)
2124.025 Million cell updates/sec

Title: US-09-825-414-7

Perfect score: 2412

Sequence: 1 MHINRVQGPVATATDSFRT.....IEGGTPASPSSEIPRPQRS 486

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt 40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	155.5	6.4	1120	1 STFR_ECOLI	P76072 escherichia
2	136	5.6	505	1 FLUB_SALTY	P52616 salmonella
3	131.5	5.5	3591	1 PHAB_BORPE	P12255 bordetella
4	131	5.4	556	1 FTTH_CLOAC	P13419 clostridium
5	129	5.3	676	1 MUCI_MESAU	O60528 mesocricetu
6	129	5.3	732	1 YAGC_ECO57	O8654 escherichia
7	128.5	5.3	1025	1 SLAF_CAUCR	P35628 caulobacter
8	127.5	5.3	401	1 YOPB_HUMAN	O06114 yersinia ps
9	127	5.3	1068	1 H1PB_HUMAN	O75146 homo sapien
10	127	5.3	1238	1 SBCC_RHOCA	O68032 rhodobacter
11	126.5	5.2	397	1 NCCB_ALCXX	O44585 alcaligenes
12	125.5	5.2	492	1 FLIC_SALRU	P06175 salmonella
13	125.5	5.2	1065	1 SEBD_YEAST	P23365 saccharomyc
14	125.5	5.2	1068	1 H1PB_MOUSE	O91455 mus musculu
15	125.5	5.2	2541	1 TAL1_MOUSE	O9490 homo sapien
16	124.5	5.2	2541	1 TAL1_MOUSE	O9490 homo sapien
17	122.5	5.1	452	1 PPOK_MYCTU	O51330 mycobacteri
18	122.5	5.1	500	1 FLUB_SALAE	P52615 salmonella
19	122	5.1	756	1 Y4S1_RHISN	P55552 rhizobium s
20	121.5	5.0	488	1 FLIC_SALCH	O06114 yersinia s
21	121	5.0	1140	1 YM9E_YEAST	O04893 saccharomyc
22	120.5	5.0	739	1 OCT1_CHICK	P15143 gallus gall
23	119	4.9	732	1 YAGC_ECOLI	P77489 escherichia
24	118.5	4.9	1714	1 SYEP_DROME	P28668 drosophila
25	118	4.9	1218	1 MGPC_MYCPN	O50341 mycoplasma
26	117	4.9	570	1 FLIR_RHOSH	O51451 rhodobacter
27	117	4.9	583	1 YP6S_MYCTU	O50733 mycobacteri
28	117	4.9	1508	1 BCSC_XANAC	P58938 xanthomonas
29	116	4.8	774	1 STF_LAMBD	P03764 bacterioph
30	116	4.8	934	1 CAPF_MYCLE	P46710 mycobacteri
31	116	4.8	1306	1 MSB2_YEAST	P32334 saccharomyc
32	115.5	4.8	397	1 DP38_MYCSM	P52851 mycobacteri
33	115.5	4.8	2249	1 OMPA_RICRI	P15921 rickettsia

34	115	4.8	342	1 GCP_HAEIN	P43764 haemophilus
35	115	4.8	559	1 FTHS_MOOTH	P21164 moorella th
36	115	4.8	802	1 BCBI_ACERY	P37716 acetobacter
37	115	4.8	1756	1 TR11_ECOLI	P14565 escherichia
38	114.5	4.7	439	1 Y579_CHLTR	O84583 chlamydia t
39	114.5	4.7	743	1 OCT1_HUMAN	P14859 homo sapien
40	114.5	4.7	920	1 MM17_MYCTU	P96289 mycobacteri
41	114.5	4.7	1783	1 RAA3_CHLRE	O9feca chlamydomon
42	114.5	4.7	2090	1 N214_HUMAN	P35658 homo sapien
43	114	4.7	575	1 FLA2_CAMJE	P22251 campylobact
44	114	4.7	1325	1 BCC3_ACERY	O94463 acetobacter
45	114	4.7	1756	1 TR12_ECOLI	P22706 escherichia

ALIGNMENTS

RESULT 1
ID STFR_ECOLI STANDARD, PRT, 1120 AA.
AC P76072; P77560;
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Side tail fiber protein homolog from lambdaoid prophage Rac.
OS STFR OR B1372.
OC Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Ohima T., Saito N.,
RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; AE000234; AAC74454.1; ALT_INIT.
DR EMBL; D90774; BAA14966.1; -.
DR EMBL; D90775; BAA14975.1; -.
DR EcoGene; EG13370; strf.
DR InterPro; IPR004089; Chmtaxis transd.
DR InterPro; IPR005003; Phage_fiber.
DR InterPro; IPR005068; Phage_fiber_2.
DR Pfam; PF03335; Phage_fiber_6.
DR Pfam; PF03406; Phage_fiber_2; 1.
KW Hypothetical protein; Fiber protein; Repeat; Complete proteome.

SQ SEQUENCE 1120 AA; 113779 MW; 542E59D71EE795B4 CRC64;

Query Match 6.4%; Score 155.5; DB 1; Length 1120;
 Best Local Similarity 23.7%; Pred. No. 0.086;
 Matches 121; Conservative 60; Mismatches 220; Indels 109; Gaps 21;

QY 19 RTASDASLASSVRSVSDQOEINAIADYLTDFVFAAHKLPPLPADSADQAAVDVHNAQI 78
 DB 115 RNASVAQNTAAKKSADASTSAREATHAAD---AASDARAASSTAGQAASQAASASS 171
 QY 79 TALITRASRLHFEGETPATIAD-TFAKAEKLDRLATTSGALRATPPAMASLLQYMPA 137
 DB 172 SA-----GTASTATATASKASAAAEKSKSAATISAGAAKTSETNASASLSQAATS 221
 QY 138 INKDWLPAPLKPPLPLISGALSGAMDQVGTQMDRATGDLHYLSASPRLDHMAA--S 195
 DB 222 ASTA-----TTKASERATSARDAASAKKSS-ETNASSASSASSATAAGNS 270
 QY 196 VK-----RHSPSLARQVLDTCVAVQT--YSARNAVRTVLAPALASRPVQGVADLG 244
 DB 271 AKAAKTSETNARSSTAAAGQSASAAAGSKTAAASASASAASTAGQAASATAAGKSAESA 330
 QY 245 VSMAGGLAANAGFNRLLSVOSRDHORGALVLGKDKPEKPAQLSEENDWLEAYKAIKSA 304
 DB 331 ASSASTATTKAGEATEQASAAARS-----ASAKTSETNAKASETS--AESKTAAS 381
 QY 305 SYSGAALNAGKRMAGLPDMATDAMGAVRSLSVSSSLTQNGLAGAGPAGVGLQEMATK 364
 DB 382 SASSAASSASS--ASASKDEATROASAKS--SATTASTKATEAAGS-----424
 QY 365 NITDPATVAASVOLTNLGASAAVAGWTAALTTPDPAVKKASFP-----IOD--TVK---414
 DB 425 -----ATAAASKST--AESATRA-----ETAARKAEDTASAVALEDASTTKGI 468
 QY 415 ---SPASSTTGVADQTVKLAKTV-----KDMGCEATHTGASLRNTVNNLR-----458
 DB 469 VOLSATNSTSETLAATPAVKASAYDNAEKLOKQONGADIPDKGCFI--NNINAVSKTDF 527
 QY 459 --QRPAREADIEEGTAASPEIIPRPMS 486
 DB 528 ADKRGMRVVRNAPAGATSGKYPPVVMRS 557

RESULT 2
 FLJB SALTY STANDARD; PRT; 505 AA.

AC P5916; P97159;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phase-2 flagellin.
 GN FLJB OR H2 OR STM2771.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 OX NCBI_TaxID=602;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=SL 375;
 RX MEDLINE=95323331; PubMed=7541401;
 RA Vanegas R.A., Jøys T.M.;
 RT "Molecular analyses of the phase-2 antigen complex 1,2... of
 Salmonella spp."
 RL J. Bacteriol. 177:3863-3864 (1995).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RL Nature 413:852-856 (2001).
 RN (3)
 RP SEQUENCE OF 1-37 FROM N.A.
 RX MEDLINE=82049491; PubMed=6271461;
 RA Silverman M., Zieg J., Mandel G., Simon M.;
 RT "Analysis of the functional components of the phase variation
 system";
 RL Cold Spring Harb. Symp. Quant. Biol. 45:17-26 (1981).
 RN (4)
 RP SEQUENCE OF 482-505 FROM N.A.
 RC STRAIN=5J2353;
 RA Minoorance J., Tanaka S., Tomimaga A., Enomoto M.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
 FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
 CC -1- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
 BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
 PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U17177; AAC3354.1; --
 DR EMBL; AE008826; AAL21657.1; --
 DR EMBL; V01370; CAA24655.1; --
 DR EMBL; D26168; BAA05156.1; --
 DR StvGene; SG10564; fljB.
 DR InterPro; IPR001492; Flagellin.
 DR InterPro; IPR001029; Flagellin.
 DR Pfam; PF00669; Flagellin_N; 1.
 DR Pfam; PF00700; Flagellin_C; 1.
 DR PRINTS; PR00207; FLAGELLIN.
 DR ProDom; PD000316; Flagellin_C; 1.
 KW Flagella; Complete proteome.
 FT INIT MET 0 0 BY SIMILARITY.
 FT CONFLICT 37 37 I -- S (IN REF. 3).
 SQ SEQUENCE 505 AA; 52404 MW; 3A5CC404AF7AF88B CRC64;

Query Match 5.6%; Score 136; DB 1; Length 505;
 Best Local Similarity 22.4%; Pred. No. 0.44;
 Matches 116; Conservative 60; Mismatches 179; Indels 164; Gaps 25;

QY 63 DSADGQAAVDVHNAQITPATIETRASRLHFEGETPATIADTFAKAEKLDRLATTSGAL-- 120
 DB 42 DDAAGQAIARFTANIKGL--TQASRNANDGIS-----IAQTTEGALNE 83
 QY 121 -----RATPFAMASL-----LQYMQPAI-----NKGDWLPAPLKPPLISGALS 162
 DB 84 INNLRQVRELAQVSANSTNSQSDLSIQAEITQRLNEIDRVSGQTQ-----FNGVKVLA 138
 QY 163 MD-----QVGTQMDRATGDLHYLSASP-----DRLHDAASVKKRHSPLARQVLDTCVAV 214
 DB 139 QDNTLTITQVANDGETTIDILKQINSQTLGDSLNVOKAYDVK-----DTAVTT 187
 QY 215 QYTSARNAVRTVLAPALASRPVQGVADLVGVSMAAGVLAANAGFNRL-----SV 264
 DB 188 KAVANNGTTLDSGLDDAIAKAAKTGTTNGTASVTGGAVKFPDADNNKYFVTIGGTGADAA 247
 QY 265 QSRDHQ-----RGGALVL---GLKDKPEK-AQLSEENDWLEAYKAIKSASYSALNAGKR 316
 DB 248 KNGDYEVNVAATDGTFTVLAAGATKTTMPAGATTTKEVQELKDTTPAVSADAKNALIAGG-- 305
 QY 317 MAGLPDMATDAMGAVRSLSVSSSLTQNGLAGAGGFA-----GVGKLOEMA 362
 DB 306 -----VD-ATDANGA--ELVQMSYTDKNGKTIEGYALKAGDKYKAAADYDEATCAIKAKT 357


```

OY      369 TK-NTDPTPKAAVSQLTLAGSAANFA-----GWT-----TALLTDP    400
DB      358 TSYTAADGTTKTAAADLGSGVDGKTEVYTIDGKTYNASKAGHDFKAQPELAERAAATTEN 417
OY      401 AVKKAESFIQ--DYKSTASSTGGVAQDTVLAKTVKDMGEAITHTGASLRNTYNLKR   458
DB      418 PLOKIDAMALAQDALRSLDGA-----YQNRFSATINLG---NITNNLS     458
OY      459 QRPAREAD-----IERGGTA--ASPSEIP 480
DB      459 EARSRIEDSDYATEVSNMSRAQIILQOAGTSVLAQAQNPV 497

RESULT 3
FHAB_BORPE
ID_FHAB_BORPE STANDARD; PRT; 3591 AA.
AC P12255;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
DE Filamentous hemagglutinin.
GN FHAB.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
CC Bordetella.
CX NCHI_Taxid=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90355839; PubMed=2388559;
RA Reiman D.A., Domenighini M., Tuomenen E., Rappunoli R., Falkow S.;
RT "Genetic characterization of Bordetella pertussis filamentous haemagglutinin: a protein processed from an unusually large precursor.";
RL Mol. Microbiol. 4:787-800(1990).
RN [2]
RP SEQUENCE OF 1-3261 FROM N.A.
RX MEDLINE=89202384; PubMed=2539596;
RA Reiman D.A., Domenighini M., Tuomenen E., Rappunoli R., Falkow S.;
RT "Filamentous hemagglutinin of Bordetella pertussis: nucleotide sequence and crucial role in adherence.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
CC -! FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND INFECTON.
CC -! SUBCELLULAR LOCATION: SURFACE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60351; AAA22974.1; .
DR EMBL; M60351; AAA22975.1; ALT INIT.
KW Antigen; Hemagglutinin.
SQ SEQUENCE 3591 AA; 367420 MW; EF7418B30DE5138 CRC64;

Query Match          5.5%; Score 131.5; DB 1; Length 3591;
Best local similarity 22.2%; Pred. No. 8.7;
Matches 111; Conservative 77; Mismatches 195; Indels 117; Gaps 24;

OY      33 SVSSPOQRIRINMADYLTHVFPAHKLPAPADBADGCAADVNNAOITLIERRA----- 86
DB      414 AASSRKDAISVNGCALXADKLSATRVV---DVDGQAVALASASSNNL-SYVAGALKKA 468
OY      87 -----SRLLPFGETPATIADTFAPAKEKDRLATTTSGALRTPFAMSLGYM-QPAIN 139
DB      469 GKLSATGRDVGVQKVAVTLGSVASDS---ALSVSAGNLRRANELVSSQLREVROGREVA 524
OY      140 KCDMLPAPLKPLTLPILSGLS-----GAMPDVC-----TKMDRATGD L 178

```

[illegible]

```
FT NP_BIND 65 72 ATP (BY SIMILARITY).
SQ SEQUENCE 556 AA; 59588 MW; 72D9D485863F0334 CRC64;

Query Match
Best Local Similarity 22.3%; Pred. No. 0.97;
Matches 111; Conservative 65; Mismatches 160; Indels 162; Gaps 26;

QY 29 SSVRSVSDQREINAIADYITDH-----VFAAHKLPPA-----DSADQQAADVHN 75
DB 3 TDIIQAOEQMKHKDVAELIDIEDDDLELYGKYKAKVSLVDVLQDKDPGKLVL----- 58
QY 76 AQITALITRASLRFEGETPATIADTAKAEKLDRLATTSSGALRATPFAMASLLQYMQ 135
DB 59 --VTAINPTAG-----EKTITNIGLSMG-----LNKIGKKTSTALR-----E 95
QY 136 PAINGDMLPAPLPLTLPLISGALSGAMDQVTKQMD---RATGDLHYLSASPDRLHDM 192
DB 96 PS-----LGPSFGVKGAGGGAQV-VPMADINLHFTGDFHAITSA-----HSL 140
QY 193 AASVSRHSPLARQVLTGAVQVTSARNAVTVLAPALASRPVQGVADLGVSMAAGLA 252
DB 141 AALVDNH-----LHGNALRIDTNRIVMKRVV--DMNDRLARKIVVGLG-GKAQGIT 189
QY 253 ANAGF-----GNRL-----SVQSRDHQRGGALVGLK 280
DB 190 REDGPDITVASEIMAILCANDREDLKERLGNMVAAYNVVDGAVRAKDLAQGALTLLK 249
QY 281 DK-----BPKAOLSEE-NDWLEAYKAISASYS-----GAALNAGK 315
DB 250 DAINPNIQVTLTENTPAFTHGPPFANIAGHCSVLTATKALKTGYAVTEAGFGADLGA 309
QY 316 -----RMAGLPDMATDAMGAVRSLVSASSLTQNGALAGFAGVKGKQEMATKNITDP 369
DB 310 PFDIKCRVAGLNPDAV-----IVA-----TVRALKMHGVA-----KEDLGTENL 350
QY 370 ATKAASVQTLNL-----AGSNAVPGAGTTAALTTPDAVKAEE-SFIQDTVKS-----TASST 420
DB 351 --DALAKGMTLHERHENAVKFPVSVVAINAFPTDTEAEKQVDFDKCKEMGVDAISDV 408
QY 421 TGYVADQTVKLAKTVKDM 438
DB 409 FAKGGGGVELAQKVIDV 426

RESULT 5
MUC1 MESAU STANDARD; PRT; 676 AA.
AC Q60528;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mucin 1 precursor.
GN MUC1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mamalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=tracheal epithelium;
RX MEDLINE=96326118; PubMed=8703480;
RA Park H., Hyun S.W., Kim K.C.;
RT "Expression of MUC1 mucin gene by hamster tracheal surface epithelial
cells in primary culture."
RL Am. J. Respir. Cell Mol. Biol. 15:237-244(1996).
CC -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC CYTOSKELETON (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U36918; AAB53965.1; -
DR InterPro; IPR00082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
KW Repeat.
FT CHAIN 1 25 POTENTIAL.
FT SIGNAL 26 676 MUCIN 1.
FT DOMAIN 26 582 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 583 603 POTENTIAL.
FT DOMAIN 604 676 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 458 573 SEA.
FT CARBOHYD 291 291 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 323 323 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 350 350 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 435 435 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 479 479 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 496 496 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 676 AA; 67616 MW; 95F479B6EC5C3884 CRC64;

Query Match
Best Local Similarity 20.6%; Pred. No. 1.6;
Matches 95; Conservative 57; Mismatches 173; Indels 136; Gaps 20;

QY 10 PPV-----TATDFRTASDASLASSVRS-----VSSDQREINAIADYITDHVFAHKL 59
DB 188 PPTSSAVNSATTPVHSGSSAPVTSSAVNSATTPVHSGSSAPVTSSAVNSATTPVHSGSSA 247
QY 60 PPADSADQAAVDVHNAQ-----ITALIETRASRLHFEGETPATIADTFAKAEKLDRLATT 115
DB 248 PPTSSVNSATTPVHSGSSAPVTSSAVNSATTPVHSGSSAPVTSSAVNSATTPVHSGSSA 293
QY 116 TSGALRATPFAMASLLQYMQ-----PAINKGDWLPAPLKPLTPLISGALSGAMDQVGT 168
DB 294 TDSA--TTPVPGSSMOTTEAIGSANTPIHNGSLVPTTSSALVPTTSAHSGASAMTNS 351
QY 169 KMDRAT-----GDL-----HYLSASPDRLHD---AQAASV-- 196
DB 352 SESDLATTPIDSGTSISTTKAPATTPVHNGSLVPTTSSVILGSSATTLIHNDTSTMAITPV 411
QY 197 -----KRH--SPSLARQVLTGVAVQVTSARNAVTVLAPALASRPVQGVADLVG 245
DB 412 GNGTQSVSPRHPVTPPPAVSSNSTIALSTYS-----TALSPAFSSHAAPQ--VSVGV 464
QY 246 SMAGGLAANAGFNGRLSVQRDQRGALVGLKQKPAQLSEENDWLEAYKAISAS 305
DB 465 SFF-----LLSPHINWQHFNSSL-----EDPSSNY-----YQELK--- 494
QY 306 YSGNALNAGKRMAGLPDMATDAMGAVRSLV--SASSLTQNGALAGFAGVAGKLV----- 358
DB 495 -----RNVSGLFLQVFSRAPFLGISTIEPRSGSVVVDSTVIFREGAVNASEVKSQ 545
QY 359 ---QEMATKNITDPATKAASVQSLTNLAGSAAVFAGWTAAAL 396
DB 546 QHEQEAEYNLAISKINVGEMQFPSSAQSPGVPFGWGIAL 586

RESULT 6
YAGR_ECO57
ID YAGR_ECO57 STANDARD; PRT; 732 AA.
AC Q8X6J4;
```

```

DR 15-UN-2002 (Rel. 41, Created)
DR 15-UN-2002 (Rel. 41, Last sequence update)
DR 15-UN-2002 (Rel. 41, Last annotation update)
DE Puerative xanthine dehydrogenase yagR, molybdenum binding subunit
DE (EC 1.1.1.204).
GN YAGR OR Z0350 OR ECS0314.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_taxid=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Poranousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T.,
RA Kumura S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.
CC -1- COFACTOR: Molybdopterin (By similarity).
CC -1- SUBUNIT: Heterotrimer of yagR, yags and yagT (Potential).
CC -1- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sb.ch).
CC -----
DR EMBL; AE005205; AAG54609.1; -
DR AP002551; BAB33737.1; -
KW Hypothetical protein; Oxidoreductase; Molybdenum; Purine metabolism;
KW Purine salvage; NAD; Complete proteome.
KW SEQUENCE 733 AA; 78135 MW; A996B91BBE2643F6 CRC64;
-----
Query Match 5.3%; Score 129; DB 1; Length 732;
Best Local Similarity 24.7%; Pred. No. 1.8; Indels 86; Gaps 21.
Matches 104; Conservative 61; Mismatches 170;
QY 22 SDASLASSSVRSVSSDQREINAIADYLDHFHAHKLPAPASADGOAVDVHNAQITL 81
DB 250 SDALLAALAAARV-----KRPVKVMLPREIPNWTTHR--PALQLHRIQAD-QSGKITVI 302
QY 82 IETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLYOWPAINKG 141
DB 303 -----SHESWGSNLLPGCGPETAVQOSEL-----LVAGNRHIGRLATL-----DLPEG 346
QY 142 DWLPAPIKELTLLIGALSGANDVOGTR-----MDRATGDLHYLSASDRLDHMAASV 196
DB 347 NAMRAPGE--AFGLM-ALEIAIDELAEKAGIDPVEFRRIINDQIDPADTR----- 394
QY 197 KRHSPLARQVLDTGAAVQTSARNAVRFVLAPALASPAVOGAVDLGVSMAG-GLAANA 255
DB 395 -RFSRQLIECLRTADKRGWKRN-----ATGQVRDGSWLVGHGVA--A 437
QY 256 GFGNLLVSQSHDHRGALVIGLKDKEKQALSEENDWLEAYKAIKSASYSGALLNACK 315

```

```

Db 438 GFRNNLL-----EKSGARV-----HLEPPIGVTVETDMD-----IGTSSYITLQATAAE 482
Oy 316 RMAGIPDPMATDAMGAIVRSLSVASSTLQNG--LALAGFAGVGLQEMATKNI-TDP-AT 371
Db 483 -MLGVPLERQVAVHLGDGSPFVSAGSGGQGMCANSTSGVAAACVLAEMTASAVGFPDEGS 541
Oy 372 KAAVVSQLTNLAGSAVAAGVMTTALTTPDAVK-----KASPIQDPTVKASTTGY 423
Db 542 QPAGCKITNGRSLIHEATAGGRLLTAESIEFGTLSKEYOSTFAGHEVEVGHASATGE 601
Oy 424 V 424
Db 602 V 602

RESULT 7
SLAP_CAUCR
ID SLAP_CAUCR STANDARD; PRT; 1025 AA.
AC P35828; Q46015; Q9REF2;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE S-layer protein (Paracrystalline surface layer protein).
GN RSHA OR CCI007.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxId=155692;
[1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=93007489; PubMed=1393820;
RA Glichrist A., Fisher J.A., Smit J.K.;
RT "Nucleotide sequence analysis of the gene encoding the Caulobacter
RL crescentus paracrystalline surface layer protein.";
RN Can. J. Microbiol. 38:193-202(1992).
[2]
RP REVISIONS TO 376; 636 AND 842-843.
RN
RP Awram P.;
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=J53001;
RA Bingle W.H., Awram P.A., Nornellint J.F., Smit J.K.;
RT "The secretion signal of C. crescentus S-layer protein is located in
RL the C-terminal 82 amino acids of the molecule.";
RN Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldlyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Elsen J., Heideberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Padke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uetehack T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Frazer C.M.;
RL "Complete genome sequence of Caulobacter crescentus.";
RN Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
[5]
RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.
RN
RP STRAIN=ATCC 19089 / CB15;
RX MEDLINE=89008089; PubMed=3049545;
RA Fisher J.A., Smit J.K., Agabian N.;
RT "Transcriptional analysis of the major surface array gene of
RL Caulobacter crescentus.";
RN J. Bacteriol. 170:4706-4713(1988).
[6]
RP CHARACTERIZATION.
RP STRAIN=ATCC 19089 / CB15;
RX MEDLINE=98292737; PubMed=9620954;

```

RA	Awram P., Smit J.K.;
RT	"The Caulobacter crescentus paracrystalline S-layer protein is secreted by an ABC transporter (type I) secretion apparatus.";
RJ	J. Bacteriol. 180:3062-3069(1998).
RL	- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. PROBABLY ACTS AS A PHYSICAL BARRIER TO PARASITES AND LYTIC ENZYMES.
CC	- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY. SECRETED BY AN ABC TRANSPORTER (TYPE I) SECRETION APPARATUS.
CC	- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	ENBL; AF062345; AAC38665.2; -
DR	ENBL; AF193063; AAF19365.1; -
DR	ENBL; AE005779; AAQ22991.1; ALT_INIT.
DR	HSSP; P22629; ISWC.
DR	TIGR; CC1007; -
DR	InterPro; IPR001343; Hemlysn_Ca_bind.
DR	Pfam; PF003353; hemolysinGabinid_3.
DR	PRINTS; PR00313; CABDNNGRPT.
KW	Cell wall; S-layer; Calcium-binding; Complete proteome.
FT	INIT MET 0
SQ	SEQUENCE 1025 AA; 98001 MW; AD7A326E1363DBAC CRC64;
	Query Match 5.3%; Score 128.5; DB 1; Length 1025;
	Best Local Similarity 22.4%; Pred.No.2.8;
	Matches 114; Conservative 57; Mismatches 175; Indels 163; Gaps 24;
QY	12 VTATDSFRTAGDASLASSVSRVSDDQOREINAIADY-----
DB	205 VSGGGVATATAAMINDLSDCALSTDNAAGVNLFTAYPSSGVSGTLSTLTGTDTLTGTA 264
QY	49 -----LTDHVFPAHKLPADSADGOAAVDVN-----AQITAL-----IETRSLRHF 91
DB	265 NNDTFVAGEAGAATLTVGDTLSGGAGTDLNVLNQAAAATPLTGVTISGIET---MNV 320
QY	92 EGEPATIAITFAKAEKLDRLATTTSGA-----LRATPMASLLQYMQPAINK 140
DB	321 TSGAAITLT-NTSSGGVTGTLANTNTTSGAAQTVTAGQNLTATTAAQAA----NNVAVDG 375
QY	141 GDWLPAFLKPLTPIISGALSAMDQVTKMDRATGDILHLYSAPDRLDMAAASVKHS 200
DB	376 G-----ANTVASTGVTSGT-TTVGAN--SAASG-----TVSVSVANS 411
QY	201 PSLARQVLDTGVAVQT--YSARNARVTVIAPALASRPVQGAVDLGVSMAGGLAANAQFG 258
DB	412 TTTTGAIATGCTAVTVQAQTAGNAVNTTLTQADVTVTCNSSTTAVTVTQTAATAAGATVA 471
QY	259 NRLLSVQSRDHQRGAULVLGKDKEPKQLSEENDWLEAYKAIKASYSYGAAALNKGMA 318
DB	472 GRV-----NGAVTI-----TDSAASATTAGK--- 493
QY	319 GLPLDMATDMCAV-RSLVSASSLTONGLAGALAGGAGVKKLQEMATKNITDPATKAASVQ 377
DB	494 -----IATVTLGSFGAAITDSALTIVNLSTGTSLSIGIR-----GALT--ATPTANTL 540
QY	378 LTNLAGSAAFVAGWTAAALTTPDPAVKKAESFDQ---TVK---STASSTTGYVADOTVKL 431
DB	541 TLNVG-----LTTTGAITDSEAADGDFTTINIAGSTASST---IASLVAAD 585
QY	432 AKTVKDMGEAIT--HTTGASLRN-TVNN 456
DB	586 ATTLNISGDARVTTISHTAAALTGITVTN 614

CC recombination reactions. The complex acts as a 3'->5' double
 CC strand exonuclease that can open hairpins. It also has a 5'
 CC single-strand endonuclease activity (By similarity).
 CC -!- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. SBC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL; AF010496; AAC16118.1; -
 DR InterPro; IPR003439; ABC transporter.
 KW Hydrolase; Nuclease; Exonuclease; Endonuclease; DNA replication;
 KW DNA recombination; ATP-binding; Coiled coil.
 FT NP_BIND 37 44 ATP (POTENTIAL).
 FT DOMAIN 395 438 COILED COIL (POTENTIAL).
 FT DOMAIN 466 487 COILED COIL (POTENTIAL).
 FT DOMAIN 521 600 COILED COIL (POTENTIAL).
 FT DOMAIN 724 770 COILED COIL (POTENTIAL).
 FT DOMAIN 901 943 COILED COIL (POTENTIAL).
 FT DOMAIN 1019 1052 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1238 AA; 128046 MW; 286BC6C63A859AE1 CRC64;

Query Match 5.3%; Score 127; DB 1; Length 1238;
 Best Local Similarity 22.6%; Pred. No. 4.4;
 Matches 131; Conservative 76; Mismatches 214; Indels 158; Gaps 28;

Qy 14 ATDSFRTASLASLSSSVRSVSSDOORINAIADYLTDFVFAHKLPPADSADGQAAVDV 73
 Db 486 AAEARDQAADSALA---KAAALQQ---ALAP---LEAAH--PPARTAD-LARIET 529
 Qy 74 HNAQTALIE--TRASRLHFEGETPATIADT---PAKAEK-----LDRLATTTSG--- 118
 Db 530 DLAELTRALRDGAEEAALGAAQAQRAATAATEALARAQESAAKQDLDRAETQIALTAP 589
 Qy 119 -----ALRA-----TPFMAASLLQYMPAINKGDWLPAPLKLPTPLISGALS 160
 Db 590 LEQADLALSDAASRLRAQLSAGSPCVCGALEHPTTFAEAGLAHLAERLRADQAAARGAAQ 649
 Qy 161 GAMOVGTGKMDRATGDHLHLSASPDRLHDMAASVKHSPSLARQVLDTGAVQVTSAR 220
 Db 650 AARDALTAQGARATAEARGTQAEDORRAQTRAEARAA-----WGDTQPRVSAR 700
 Qy 221 NAVRTVLAPALASRP---AVQAVD---LGVSMAGGLAANAGFGNRLLSVQSRDHQGG 273
 Db 701 -----PLAPALPGTDPPTALAAQDRLCALQTAEEAAQAEISALRUTLTAEB-RDRERKL 754
 Qy 274 ALVL---GLKDK--EPKQALSEENDWLEAYKAIKASYSYGAAAL-----NAGKRMAGLP 321
 Db 755 RALLAHRGTRERLAVQOETAQEAALAEARETEAAARDGLALALAPALARAAGEDDPAAP 814
 Qy 322 -----LDWATDAMGAVRSLVSASSLTQNGL-ALAGGFGAGVGLQEMATKNITDPAKAA-- 373
 Db 815 GLAERLAATVSAVGAARTGLQAA---QEALSAALPQLAAARRDSETATAQAQSAQAARD 871
 Qy 374 ---AVSQTNLNAGSAVAGWTTA-----ALTTDPAVKKAS 407
 Db 872 RDGAAALR--AERAPLDGQPTALHRFRFDORLAAQRAQAAADLATAQAALAAEA 929
 Qy 408 FIQDTVKSTASSTGY-----VADQTVKLAKTV-----KDMGG--E 441
 Db 930 RAAETARAASEAATAQRAAEADLAAALEAAASAEALALALPPGTAERLRADLRGRDD 989
 Qy 442 AITHTGASLRVNNLRORPAREADIEGGGTAAPSEIYP 480
 Db 990 AVTAARSAL-----GARKTDLQIEDQGNPAEDP 1018

RESULT 11

NCBB ALCXX
 ID NCBB_ALCXX STANDARD; PRT; 397 AA.
 AC Q44585;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Nickel-cobalt-cadmium resistance protein nccb.
 GN NCCB.
 OS Alkaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).
 OG Plasmid pTOM9.
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 OC Achromobacter.
 OX NCB1_TaxID=515;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=31A;
 RX MEDLINE=95050278; PubMed=7961470;
 RA Schmidt T., Schlegel H.G.;
 RT "Combined nickel-cobalt-cadmium resistance encoded by the ncc locus
 of Alkaligenes xylosoxydans 31A.";
 RL J. Bacteriol. 176:7045-7054(1994).
 CC -!- FUNCTION: COMPONENT OF THE NCC CATION-EFFLUX SYSTEM THAT CONFERS
 CC RESISTANCE TO NICKEL, COBALT AND CADMIUM.
 CC -!- SIMILARITY: HIGH, TO A. EUTROPHUS CNRB AND SOME, TO ALCALIGENES
 CC CZCB.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

Query Match 5.2%; Score 126.5; DB 1; Length 397;
 Best Local Similarity 23.2%; Pred. No. 1.2; Indels 79; Gaps 18;
 Matches 88; Conservative 59; Mismatches 153;
 Qy 149 KPLTPLISGALS-----GAMDOVGTGKMDRATGDHLHLSASPDRLHDMAASVKRHSP 201
 Db 9 QPSWPIAGVAAALVGFCAAGLGSP-----SGAEVSKLAAPEK-----AASAPAAEP 60
 Qy 202 SLAR---QVLDTGAVQVTSARNAVTVLAPALASRPVAGVDLGVSMAGGLAANAGF 257
 Db 61 AEVRIPGEYLAANIAVEPVSAAGVGSVLLAP--ASVAAVPGSEAVIASRAAG----- 111
 Qy 258 GNRLLSVQSR--DHORGALVGLKQKEPKAQLSEENDWLEA-----YKAIKSASYS 307
 Db 112 --AVLRIQKGLGDAVRAGD-VLALVDSPEAAAMAAERKVAQARADLARKTYERESSLPQ 168
 Qy 308 GAALNAGKRMAGLPDMDATDAMGAVRSLVSASSLTQNG--LALAGGFGAGVGLQEMATKN 365
 Db 169 GVTPROEMESARIALDVAAQAEVQEAATVAQAQKSVSSDGRSVAVVSPAG-----RITAQS 223
 Qy 366 ITDPATKAAVSQTLNAGSAV--FAGWTTAALT-----TDPVKKAESF-IQTVKSTA 417
 Db 224 VTLGAYVAPQAEELFRVAGSGAVQVEAYVTAAATSRIAAGSDATITVLANGAPLAGRQAVT 283
 Qy 418 SSTTGYVADQTVKLAKTVKDMG--GEAI---THTGASLRN-----TNNLRQRPARE 464
 Db 284 PTVSGSARAATVVTVFDVANSGLIVGEGVQVRLHTKAADANAMSVPEDAVQNLDRGVVF 343
 Qy 465 ADIEGGTAAPSEIYP 483
 Db 344 VRTOQG-----FRP 352

```

RESULT 12
FLIC_SALRU STANDARD; PRT; 492 AA.
ID ID FLIC_SALRU STANDARD; PRT; 492 AA.
AC P06175;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FLAGELLIN (Phase-1-R flagellin).
GN FLIC OR FLAP OR HAG.
OS Salmonella rubislaw.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OC NCBI_TaxID=598;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10717;
RX MEDLINE=87040788; PubMed=3774559;
RA Wei L.-N., Jøys T.M.;
RT "The nucleotide sequence of the H-1r gene of Salmonella rubislaw.";
RL Nucleic Acids Res. 14:8227-8227(1986).
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -1- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
CC BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
CC PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL, X04505; CAA8190.1; -.
DR PIR, S07375; S07375.
DR InterPro; IPR001492; Flagellin_N.
DR InterPro; IPR001029; Flagellin_C.
DR Pfam; PF00669; Flagellin_N; 1.
DR Pfam; PF00700; Flagellin_C; 1.
DR PRINTS; PR00207; Flagellin.
DR ProDom; PD000316; Flagellin_C; 1.
DR FLAGELLIN.
DR KW Flagella.
DR FT INIT MET.
SQ SEQUENCE 492 AA; 51285 MW; 0F87CB00D311B5FA CRC64;
Query Match 5.2%; Score 125.5; DB 1; Length 492;
Best Local Similarity 21.9%; Pred. No. 1.7;
Matches 110; Conservative 56; Mismatches 192; Indels 145; Gaps 20;
QY 63 DSADGQAAVNVNNAQTALLETTRSRHFEGETPATTAADPAFAKEDRLATTGSL-- 120
DB 42 DDAAGQAIARFTANIGL--TQASRNANDGIS-----IAQTTEGALINE 83
QY 121 -----RATPFAMASL-----LOYMOPALINK-----GDMFAPAPKPL 151
DB 84 INNNGQVRELAIVOSANSTNSQSLDSIOAEITQRLNEIDRVSGQTQPNQKVALQONTL 143
QY 152 TPILSALSGAMDOVGTRKMDRATGDLHYLSASPDRLHDMMAASVKRHSPLARQVLDTG 211
DB 144 TIQV-GANDETIDIDKQINSQTLGDLTLNVQOKYKVSPTAAAVTGTDSA-----TA 196
QY 212 VAVQTVSARAAVRTVLAAPALASRAVQAVDLG-----VSMAGGLAANFGNRLISVQ 265
DB 197 IDKSTFPAAS-----ATTIGGTPAIRIGDKLPDDTGTGYVADVSGTAKGVVEVTAAD 249
QY 266 SRDHQRG---GALVYLGKD---KEPKAQLSBENDMLEAVYKAKSASGALNAGKRMAG 319
DB 250 GXYTLTCTPTGPTAGPPTATKDVKOTQOENADLTAKAKALTRAGVAAA----- 299
QY 320 LPIDMAIDMAGAVRSIVSASSLTQNGIALAGFA-----GVGKLDEMATKNI 366

```

```

DB 300 -----GHRSVYKMSYTDNNNGKTIIDGLAVKVDYYSATQNKDGSISINTTKYT 348
QY 367 TDPAT-KAAYSQTLNLNLAGSAV---PAGWTTAALTTPDAVKKASFIQDTYKSTASSTGY 423
DB 349 ADNGTSKTLANKLGGADGKTEVVSIGGKTYAASKABSHNFKAQ---PDLEAAATTEMP 405
QY 424 VADQTVLAK-----TVKDMGGEAITHTGASLERNVNNLRQRPAREAD----- 466
DB 406 LQKIDALAQVDTLRSDLGAVQGRFNSAITNLG-----NTVNNLTASRSRLSDSYATEVS 461
QY 467 -----IEEGGTA--ASPEIP 480
DB 462 NMSRAQITLQAGTSVLAQANQVP 484
RESULT 13
SED4_YEAST STANDARD; PRT; 1065 AA.
ID SED4_YEAST STANDARD; PRT; 1065 AA.
AC P25365;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE SED4 protein.
GN SED4 OR YCR067C OR YCR67C OR YCR901.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OC NCBI_TaxID=4932;
OX [1]
RN RN
RP SEQUENCE OF 1-514 FROM N.A.
RA Contreras R., Demolder J., Fiers W., Molemans F.;
RT Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 446-1065 FROM N.A.
RX MEDLINE=92221691; PubMed=1561837;
RA Bent P., Chanet R., Fabre G., Fukuhara H., Sor F.;
RT "Sequence of the sup61-RAD18 region on chromosome III of
RT Saccharomyces cerevisiae.";
RL Yeast 8:147-153(1992).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=93011014; PubMed=1327759;
RA Hardwick K.G., Boothroyd J.C., Rudner A.D., Pelham H.R.B.;
RT "Genes that allow yeast cells to grow in the absence of the HDEL
RT receptor.";
RL EMBO J. 11:4187-4195(1992).
CC -1- FUNCTION: WHEN PRESENT IN MULTIPLE COPIES, ALLOWS CELLS TO GROW
CC IN THE ABSENCE OF HDEL RECEPTOR BY EITHER SLOWING TRANSPORT FROM
CC THE ER OR BY STIMULATING VESICLE BUDDING FROM GOLGI MEMBRANES.
CC -1- SUBCELLULAR LOCATION: TYPE II membrane protein.
CC -1- MISCELLANEOUS: IN THE PROCESS OF TRANSPORT, SED4 ITSELF MAY
CC MIGRATE TO THE GOLGI APPARATUS AND FUNCTION IN SUBSEQUENT
CC TRANSPORT EVENTS.
CC -1- SIMILARITY: CONTAINS 2 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: IN THE N-TERMINAL HALF, TO YEAST SEC12.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL, X59720; CAA42273.1; -.
DR PIR, S19482; S19482.
DR SGD; S0000663; SED4.
DR InterPro; IPR000886; ER_target.
DR InterPro; IPR001860; WD40.
DR Pfam; PF00400; WD40; 2.
DR SMART; SM00320; WD40; 2.
DR PROSITE; PS00014; ER_TARGET; 1.

```

```
DR PROSITE; PS00678; WD REPEATS 1; FALSE NEG.
DR PROSITE; PS0082; WD REPEATS 2; FALSE NEG.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW Transport; Protein transport; Glycoprotein; Endoplasmic reticulum;
KW Transmembrane; Signal-anchor; Repeat; WD repeat.
FT DOMAIN 1 346 CYTOPLASMIC (POTENTIAL).
FT TRANSEM 347 365 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 366 1065 LUMENAL (POTENTIAL).
FT REPEAT 259 298 WD 1.
FT REPEAT 302 341 WD 2.
FT DOMAIN 467 476 POLY-SER.
FT DOMAIN 579 590 POLY-SER.
FT DOMAIN 824 863 4 X 10 AA TANDEM REPEATS.
FT REPEAT 824 833 1.
FT REPEAT 834 843 2.
FT REPEAT 844 853 3.
FT REPEAT 854 863 4.
FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1039 1039 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 1062 1065 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 1065 AA; 114079 MW; 0FIDAE0D428085A8 CRC64;

Query Match 5.2%; Score 125.5; DB 1; Length 1065;
Best Local Similarity 19.8%; Pred. No. 4.5;
Matches 90; Conservative 93; Mismatches 171; Indels 101; Gaps 22;

QY 12 VTATDSFTASDASLASSSVSSDQOREINATADYLDTHVFAHKLPPADSDAQAAV 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 585 LSLSSLTSPPTALTSTATVTTQNTDAN--TSFL-----DNSKPASTR 633

QY 72 DVHNAQITAIETRASRLHFEGETPATIADTFKAELDLRLATTSGALRATPP-AMASL 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 634 EIIYKTKIITEVITKIEYRNI-----PA--SDGNAEAEQV---VTTSSMLLTPTDWTWS- 682

QY 131 LQYMPAINKGDWLPAPLKPL--TFLISGALSGAMDQVGTKMDRATGDLHLVLSAPDRL 188
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 683 -----SPVSEIDPIASELERVMTETHTSISASEFDSVASNLIPNE---EILSTASQD-- 733

QY 189 HDMAAASVKRHSPLARQVLDTGAVQVYSARNVTVLAPALASR--PAVQG----- 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 734 -----SISHPSTSDSITSG--FQIE---VSTVSSVLASESPISDSFSSKFH 781

QY 240 AVDLGVSMAGGLAANAGFG-----NRLLSVQSRHQGGALVLGLKDKPKAQLSEND 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 782 SISEPVSSAIVETATSFSTETKTSRVIATFSTEDSERSALI-----DNSEYTSVLADN- 836

QY 294 WLEAYKAIKSASYSGAALNAGKRMAGLPIDMATDAMGAVRSILVSASSLTQN----- 344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 837 -LEPTSVLADNSEPTSVLADSE-----PTSVPTDVAQPKTSVGGSSLSSESTNIEGTSMA 891

QY 345 GLALAGGAGVGKLOEMATKNITDPATKAASVQLTNLAGSAAVFAGWTAAALTTDPAVKK 404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 892 SMIFSSSGASIGALSDIGKGLSVESASSTVAQ-----PMSGVTTT----- 932

QY 405 AESFTQDTVKSTASS--TTGVVADQTVKLATVKD 437
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 933 APSFVSSPHKISASSIDASGFVQKEIMIEVQSSKD 967

RESULT 14
HPR MOUSE
ID_HPR_MOUSE STANDARD; PRT; 1068 AA.
AC Q9UKY5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Huntingtin interacting protein 1 related (Hipl-related).
GN HPI1R.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
```

```
[1]
SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=20082888; PubMed=10613908;
RA Engqvist-Goldstein A.E.Y., Kessels M.M., Chopra V.S., Hayden M.R.,
RA Drubin D.G.;
RT "An actin-binding protein of the Sla2/Huntingtin interacting protein 1
RL J. Cell Biol. 147:1503-1518(1999).
CC -1- FUNCTION: COMPONENT OF CLATHRIN-COATED PITS AND VESICLES, THAT MAY
CC LINK THE ENDOCYTIC MACHINERY TO THE ACTIN CYTOSKELETON.
CC -1- SUBUNIT: INTERACTS WITH ACTIN AND HUNTINGTIN INTERACTING PROTEIN 1
CC (HPI). DOES NOT INTERACT WITH HUNTINGTIN (BY SIMILARITY)
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. MEMBRANE-ASSOCIATED PROTEIN,
CC MAINLY LOCALIZED AT THE ENDOCYTIC COMPARTMENTS AND IN THE
CC PERINUCLEAR REGION.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. EXPRESSED AT LOWER LEVELS IN
CC SKELETAL MUSCLE AND HEART. THE LEVEL OF EXPRESSION DOES NOT CHANGE
CC APPRECIABLY DURING DEVELOPMENT.
CC -1- DOMAIN: THE TALIN-LIKE DOMAIN BINDS F-ACTIN.
CC -1- SIMILARITY: BELONGS TO THE SLA2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; AF221713; AAF34662.1; --
DR MGD; MGI:1352504; Hiplr.
DR InterPro; IPR001026; ENTH.
DR InterPro; IPR002558; ILWEQ.
DR Pfam; PF01417; ENTH; 1.
DR ProDom; PD01608; ILWEQ; 1.
DR ProDom; PD0118207; ILWEQ; 1.
DR SMART; SM00273; ENTH; 1.
DR SMART; SM00307; ILWEQ; 1.
KW Actin-binding; Coiled coil.
FT DOMAIN 77 80 POLY-SER.
FT DOMAIN 311 316 POLY-GLU.
FT DOMAIN 346 644 COILED COIL (POTENTIAL).
FT DOMAIN 800 1068 TALIN-LIKE.
SQ SEQUENCE 1068 AA; 119484 MW; 97CEB9D92CDF5DB1 CRC64;

Query Match 5.2%; Score 125.5; DB 1; Length 1068;
Best Local Similarity 20.9%; Pred. No. 4.5;
Matches 117; Conservative 72; Mismatches 206; Indels 165; Gaps 24;

QY 19 RTASDASLASSSVRSVSSDQOREINATADYLDTHVFAHKLPPADSDAQAAVDVHNAQI 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 537 RTEQSGSELSSRLDTLNAEKE---ALSGVVRQ-----EAELAAQSLVREKE 581

QY 79 TALETRASRLHFEGETPATIADTFKAEL-----DRLATTSGALRATPPAMASILQ 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 582 EALSQEQQRSSQKGEIURLQAEKESQEQGLRQKLLDEQLAVLSAAAEAEAI----- 634

QY 133 YMQPAINKGDWLPAPLKPL-----TP--LISGALSGAMDQVGTKMDRATGDLHLVLSAS 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 635 -LQDAVSKLD-----DPLHLRCTSSPDYLVSR-QAALDSVS---GLEGHTQYLASS 682

QY 185 PDLRHDAASVKRHSPLARQVLDTGAVQVYSARNV----- 223
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 683 EDA--SALVAALTRFS--HLAADTVNGAATSHLAPTDPADRLMDTCRECGARALELVQL 739

QY 224 --RVLAPALAS--RPVQGAVIDLVGSMAGGLAANAGFNLSSVQSRDHQGGALVLGL 279
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 740 QDQTVLRRQFSLMRAPLQGLQLGQDL-----KPKSLDVQRBEIG-----AM 782

QY 280 KDEPKFQKLGSENDWLEAYKAIKSAS---YSGAALNAGKRMAGLPIDMATDAMGAVRSLV 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 783 VDKEMWATSAIEDAVRRIEDMMSQARHSSGVKLEVNERI-----LNSCTDLMKAILLV 838
```


QY 337 SASLTONGLALAGFAGVGLQEMATKNITDPATKAASQLTNIAGSAVACWTAL 396
 DB 839 MTSTSLQKEIVESG--RGAATQGEFYAKN-----SMTBGLSASAVCMGATQL 886
 QY 397 --TTDPAVKRAESTIQTVKS--TASSTGVADQYK-----LATYVDKG 439
 DB 887 VESADKVLHMKYELIVCSHEIJAASQALVAASKYKANKNSPHLSRLQECSTVNERA 946
 QY 440 GEATHT-----GASLRNTVNNLRORPARADIE----- 469
 DB 947 ANYVASTKSGEQIEDDTMDPFGSLSLIKKQEMETQVLELEKTLAEARVNLGELRK 1006
 QY 470 -----GTAASPSRIFPR 483
 DB 1007 OHVLAGMGTPSEEPSRP 1026

RESULT 15
 TALI MOUSE
 ID TALI MOUSE STANDARD; PRT; 2541 AA.
 AC P26039;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE TALI.
 GN TALI OR TLN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Fibroblast;
 RX MEDLINE=91015390; Pubmed=2120593;
 RA Rees D.J.G., Adees S.A., Singer S.J., Hynes R.O.;
 RL "Sequence and domain structure of talin.";
 Nature 347:685-689(1990).
 CC -1- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL
 STRUCTURES TO THE PLASMA MEMBRANE. HIGH MOLECULAR WEIGHT
 CYTOSKELETAL PROTEIN CONCENTRATED AT REGIONS OF CELL-SUBSTRATUM
 CONTACT AND, IN LYMPHOCYTES, AT CELL-CELL CONTACTS.
 CC -1- SUBUNIT: BINDS WITH HIGH AFFINITY TO VINCLIN AND WITH LOW
 AFFINITY TO INTEGRINS.
 CC -1- PTM: PHOSPHORYLATED.
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -1- SIMILARITY: IN THE C-TERMINUS TO YEAST SLA2 AND C. ELEGANS ZK370.3.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X56123; CA39588.1; -
 CC PIR; S11661; S11661.
 CC MGD; MGI:1099832; Tln.
 CC InterPro; IPR000299; Band_4.1.
 CC InterPro; IPR002558; ILMEQ.
 CC InterPro; IPR000159; RA domain.
 CC Pfam; PF00373; Band_4.1; 1.
 CC Pfam; PF01608; ILMEQ; 1.
 CC ProDom; PD011820; ILMEQ; 1.
 CC SMART; SM00295; B4.1; 1.
 CC SMART; SM00307; ILMEQ; 1.
 CC PROSITE; PS00660; BAND_4.1; 1.
 CC PROSITE; PS00661; BAND_4.1.2; 1.
 CC PROSITE; PS00664; BAND_4.1.3; 1.
 CC Structural protein; Cytokeleton; Phosphorylation.
 FT DOMAIN 173 332 BAND 4.1-LIKE.
 FT VARIANT 1105 1105 L -> P.

FT VARIANT 2180 2180 K -> M.
 SQ SEQUENCE 2541 AA; 269832 MW; 14EF75ABE9FC2CBB CRC64;
 Query Match 5.2%; Score 125.5; DB 1; Length 2541;
 Best Local Similarity 22.2%; Pred. No. 13;
 Matches 129; Conservative 69; Mismatches 242; Indels 141; Gaps 26;
 QY 8 QQPP---VTATDSFRTASDA--SLASSSVRSVSDQ--QREINADYLT-----HVFA 55
 DB 1625 RDPFRMVLGHG--RTVSDSIKKLITSMRDKAPQGLECFETALNLSCLDLDOASLAAY 1683
 QY 56 AHKLPPADSADQQAADVHNAQITTA-----LIETASRLHFGE-----TP 96
 DB 1684 SQQLAPREGISOBA--LHTQMLTAVOEISHLIEPLASARASASQUGHRVSOQAQYFEP 1740
 QY 97 ATTADTFKAER-----LDRLATTTSGALRA-----TPRMSL 131
 DB 1741 LTLAAGAAKSTLSHPQOMALLDQTKTLASALQULTYTAKEAGNPKQAHTQDEABAV 1800
 QY 132 QYQPAINKGDMPLAPLKPPLTPLISGALSGAMPDQVGTQMDRATGDLHYLSASPRLHDA 191
 DB 1801 QMTTEAV-----EDLTTLNEASNA--GVGGMWDSITQAINQDEGP--MGDP 1846
 QY 192 MAASVKRHSPLARQVLDTCVAVQTSAR--NAVRTVLAPLASRPVQAVDLGVSMAGG 250
 DB 1847 EGSEFVD--YQTTMVRTAKAIVTVQEMVTKSNTSPBEIGPLANOLTSYGRLASQAKPAAY 1905
 QY 251 LAANAGFGNLL--SYGSRDH-----QRGALVGLDKXEPKQOLSRENMLEAYKAIS 303
 DB 1906 AAENEIEGAHITKRRVOELHGCSALVTYKAGALQCSPSDYTKKELLE----CRRVSEK 1960
 QY 304 ASYSGAALNAGK-----MAGLPIDMATDMAGAVRSIVSASLTONGIALAG-- 350
 DB 1961 VSHVLAALQGNNGTQACITRAAVSGIIDLDTTIFA-----TAGTLRBEGETPADH 2015
 QY 351 --GPAGVGLQEMATKNITD-----PATKAASQLTNLA-----GSAVFAWTT 393
 DB 2016 REGILTKAKVLVEDTKVLVQNAASQOEKLAQAASVATITRLADVVKLGAASISGAEDPE 2075
 QY 394 AALTTDAVVKKASFIQDTYKSTASSTGVAD-----QTVLAKTVKD 437
 DB 2076 TQVVLINAVDVAKALGDLISAT--KAAAGVGGDDPAVMQKNSAKVWTVNTSLTKVKA 2134
 QY 438 MGGEATHTGASLRNTVNNLRORPARADIEEGTAASPE 478
 DB 2135 VEDEATGTGA--LEATTEHROELAVCSPPEPAKSTPED 2174

Search completed: January 2, 2003, 15:19:19
 Job time : 16.4902 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:17:36 ; Search time 13.4861 Seconds
(without alignments)
1060.315 Million cell updates/sec

Title: US-09-825-414-7

Sequence: 1 MHINRRVQPPVATDSEFT.....IEEGTASPSSEIRPFRMRS 486

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/prodata/1/aa/55-COMB.pep.*
2: /cgn2_6/prodata/1/aa/56-COMB.pep.*
3: /cgn2_6/prodata/1/aa/56-COMB.pep.*
4: /cgn2_6/prodata/1/aa/56-COMB.pep.*
5: /cgn2_6/prodata/1/aa/PCITUS-COMB.pep.*
6: /cgn2_6/prodata/1/aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	130.5	5.4	1996	US-08-804-227C-9	Sequence 9, App11
2	130.5	5.4	1996	US-08-804-198-3	Sequence 3, App11
3	128.5	5.3	1026	US-08-194-290-7	Sequence 7, App11
4	127.5	5.3	401	US-08-591-079-6	Sequence 6, App11
5	127	5.3	676	US-09-085-1998-7	Sequence 7, App11
6	127	5.3	2123	US-08-968-685A-10	Sequence 10, App1
7	126.5	5.2	669	US-09-071-035-264	Sequence 264, App
8	126.5	5.2	1026	US-08-614-377A-7	Sequence 7, App11
9	126.5	5.2	1026	US-09-142-648B-7	Sequence 7, App11
10	126.5	5.2	1638	US-09-071-035-258	Sequence 258, App
11	126.5	5.2	1638	US-09-071-035-262	Sequence 262, App
12	126.5	5.2	1638	US-09-071-035-266	Sequence 266, App
13	125.5	5.2	1068	US-09-085-1998-11	Sequence 11, App1
14	123	5.1	1024	US-09-562-737-41	Sequence 41, App1
15	122	5.1	488	US-08-809-326A-1	Sequence 1, App1
16	122	5.1	649	US-08-809-326A-15	Sequence 15, App1
17	122	5.1	5215	US-09-105-537-2	Sequence 2, App1
18	118.5	4.9	391	US-08-818-112-107	Sequence 107, App
19	118.5	4.9	391	US-08-818-111-107	Sequence 107, App
20	118.5	4.9	391	US-09-056-556-107	Sequence 102, App
21	118.5	4.9	1024	US-09-072-596-102	Sequence 102, App
22	118.5	4.9	1024	US-09-562-737-44	Sequence 44, App1
23	118	4.9	2285	US-09-308-375-2	Sequence 2, App1
24	117.5	4.9	1024	US-09-562-737-43	Sequence 43, App1
25	116	4.8	1024	US-09-562-737-45	Sequence 45, App1
26	115	4.8	1832	US-09-335-409-4	Sequence 4, App1
27	115	4.8	1832	US-09-568-102-4	Sequence 4, App1

28	115	4.8	1832	US-09-567-969-4	Sequence 4, App11
29	115	4.8	1832	US-09-568-480-4	Sequence 4, App11
30	115	4.8	1832	US-09-568-486-4	Sequence 4, App11
31	115	4.8	1832	US-09-568-472-4	Sequence 4, App11
32	115	4.8	1832	US-09-567-899-4	Sequence 4, App11
33	115	4.8	3031	US-07-689-008-2	Sequence 2, App11
34	114.5	4.7	748	US-09-066-046-4	Sequence 4, App11
35	114.5	4.7	2756	US-08-375-709-11	Sequence 11, App1
36	114.5	4.7	2756	US-08-752-929-11	Sequence 11, App1
37	114.5	4.7	2756	US-09-090-793-7	Sequence 7, App1
38	114	4.7	688	US-09-141-047-8	Sequence 8, App11
39	114	4.7	7257	US-09-335-409-5	Sequence 5, App11
40	114	4.7	7257	US-09-568-102-5	Sequence 5, App11
41	114	4.7	7257	US-09-567-969-5	Sequence 5, App11
42	114	4.7	7257	US-09-568-480-5	Sequence 5, App11
43	114	4.7	7257	US-09-568-486-5	Sequence 5, App11
44	114	4.7	7257	US-09-568-472-5	Sequence 5, App11
45	114	4.7	7257	US-09-567-899-5	Sequence 5, App11

ALIGNMENTS

RESULT 1
US-08-804-227C-9
; Sequence 9, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kunzsch, Stuart A.
; APPLICANT: Roelck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1996 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-804-227C-9
Query Match 5.4%; Score 130.5; DB 2; Length 1996;
Best Local Similarity 21.5%; Pred. No. 0.008;
Matches 133; Conservative 70; Mismatches 216; Indels 199; Gaps 28;
QY 12 VFTDPSRTASDASLSSSVRSVSQDQREINAIYLDHV-----FAH 57
DB 1342 LTTTAAGRITSSDGG-RISLSDDEVATVRRLEAVQARLADTETAPDVDTAARPTTAA 1400
QY 58 KLPDASADQAAVDVHNAQITALETRASRLHFEQ--ETPATIA---DTFAAKELDR 111

Db 1401 TSPRVDTTGARTAD--GPRLVLTGAGP---EGGAADPAGAAWGLVVRVAQAEQGR 1455
Qy 112 LATT-----TSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPPLI----- 155
Db 1456 FTLVDVDTGQASRLPGLLAT--DAGQSAVRDGRVTVPRVLPVADVPVPHGGGTAADGTG 1513
Qy 156 SGALSGANDQVGTQMDRATGDLHYLSASPRDLHDAASVKRH----- 199
Db 1514 AGEPSATLDPEGTVLITGGTG-----ALAAETARHLVDRHKVRHLLVGR 1559
Qy 200 ---SPSLARQVLD-----TGVAVQT--YSARNVTVLAPALASRPVQGVADVLGVSNA 249
Db 1560 GPDAPGVDRVLVAELTESGAEVAVRACDVTDRDALRLLDLPDDEHPLTCVVHTAGVLD 1619
Qy 250 GLAANAGFNRLLSVQSRDHQGGALVLGKDKPKAQLSEENDWLE-----AYKA 300
Db 1620 VLSAQT--AERIDTV-----LRPKADAHVHLDLREIGRVPLVLYSS 1660
Qy 301 IKS-----ASYSGA-----ALNAGKRMAGLP-----LDMATDAMGAVRSLVS 337
Db 1661 VSATLGSAGQAGYAANAFMDALAARRCAAGHPALSGLGWMSGVGLATGLDGA-----D 1715
Qy 338 ASSLTONGLALAGGAGVGVKLGQEMATKNITDPATKAAYSQTLNLAGSAV-----FAGW 391
Db 1716 AARVRSGLAPLDAGAALDLDRLA--PEPALLPVRLDRAAGATLPEVLRDLAGV 1773
Qy 392 TTAALT-----DPAVKA-----ESFIQDTVKSTASS 419
Db 1774 PADARSTPGAAGTGDEGAVRPAPAPADAAGTLAARLAGRSAPERTALLDLVRTEVAA 1833
Qy 420 TTGVVADQTVKLAKTVKMGGEALTHTGASLRNTVN-----NLR-----QRPAREAD- 466
Db 1834 VLGHGDPAAIGAARTFKDAGFDSL--TAVDLNRNLNRTGLRLPATLVFDHPTPLALAE 1891
Qy 467 ---IEEGGTAASPEIP 480
Db 1892 LLDGLEAAGPAEPAAEVP 1909

RESULT 2

US-08-804-198-3
; Sequence 3, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rosteck, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1996 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-804-198-3

Query Match 5.4%; Score 130.5; DB 2; Length 1996;
Best Local Similarity 21.5%; Pred. No. 0.008;
Matches 133; Conservative 70; Mismatches 216; Indels 199; Gaps 28;

Qy 12 VTATDSRRTASDASLASSVRSVSSDQOREINAIADYLDHV-----FAAH 57
Db 1342 LTTAAAGRTASDGG-RLSLDEVRATVRTRLEAVQARLADTETAPDVRVTAARPTAAR 1400
Qy 58 KLPPADSADGQAAVDVHNAQITALITETASRLHPEG--ETPATIA---DTFAKAELDR 111
Db 1401 TSPRVDTTGARTAD--GPRLVLTGAGP---EGGAADPAGAAWGLVVRVAQAEQGR 1455
Qy 112 LATT-----TSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPPLI----- 155
Db 1456 FTLVDVDTGQASRLPGLLAT--DAGQSAVRDGRVTVPRVLPVADVPVPHGGGTAADGTG 1513
Qy 156 SGALSGANDQVGTQMDRATGDLHYLSASPRDLHDAASVKRH----- 199
Db 1514 AGEPSATLDPEGTVLITGGTG-----ALAAETARHLVDRHKVRHLLVGR 1559
Qy 200 ---SPSLARQVLD-----TGVAVQT--YSARNVTVLAPALASRPVQGVADVLGVSNA 249
Db 1560 GPDAPGVDRVLVAELTESGAEVAVRACDVTDRDALRLLDLPDDEHPLTCVVHTAGVLD 1619
Qy 250 GLAANAGFNRLLSVQSRDHQGGALVLGKDKPKAQLSEENDWLE-----AYKA 300
Db 1620 VLSAQT--AERIDTV-----LRPKADAHVHLDLREIGRVPLVLYSS 1660
Qy 301 IKS-----ASYSGA-----ALNAGKRMAGLP-----LDMATDAMGAVRSLVS 337
Db 1661 VSATLGSAGQAGYAANAFMDALAARRCAAGHPALSGLGWMSGVGLATGLDGA-----D 1715
Qy 338 ASSLTONGLALAGGAGVGVKLGQEMATKNITDPATKAAYSQTLNLAGSAV-----FAGW 391
Db 1716 AARVRSGLAPLDAGAALDLDRLA--PEPALLPVRLDRAAGATLPEVLRDLAGV 1773
Qy 392 TTAALT-----DPAVKA-----ESFIQDTVKSTASS 419
Db 1774 PADARSTPGAAGTGDEGAVRPAPAPADAAGTLAARLAGRSAPERTALLDLVRTEVAA 1833
Qy 420 TTGVVADQTVKLAKTVKMGGEALTHTGASLRNTVN-----NLR-----QRPAREAD- 466
Db 1834 VLGHGDPAAIGAARTFKDAGFDSL--TAVDLNRNLNRTGLRLPATLVFDHPTPLALAE 1891
Qy 467 ---IEEGGTAASPEIP 480
Db 1892 LLDGLEAAGPAEPAAEVP 1909

RESULT 3

US-08-194-290-7
; Sequence 7, Application US/08194290
; Patent No. 5500353
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H
; TITLE OF INVENTION: Bacterial surface protein expression
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shlesinger, Arkwright & Garvey
; STREET: 3000 South Eads Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA

```

; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,290
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Garvey, George A
; REGISTRATION NUMBER: 17737
; REFERENCE/DOCKET NUMBER: 5946-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-684-5600
; TELEFAX: 703-836-5288
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1026 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-194-290-7

```

```

Query Match      5.3%; Score 128.5; DB 1; Length 1026;
Best Local Similarity 22.4%; Pred. No. 0.0043;
Matches 114; Conservative 57; Mismatches 175; Indels 163; Gaps 24;

```

```

QY 12 VTATDSRTASDASLASSSVRSVSSDOOREINAIADY----- 48
DB 206 VSGIGAYTAAMINDLSGALSTDAAGVNLPTAPVSSGSGSTLSTGTDTLTGTA 265
QY 49 -----LTDHPAAHKLPPADSADGAADVAN-----AQITL-----IETRSLH 91
DB 266 NNDTFVAGEVAGATLVGDTLSSGACTDVUNVQAAVTLPTGVTISGIE-----MNV 321
QY 92 EGETPATIADTFAKAEKLDRLATTSGA-----LRATPFAMASLLQVNPAPINK 140
DB 322 TSGAAILTLNTSGSVTGLTALNTVTSAGAQVTAGAGQNLATTAQA-----NNVAVDG 376
QY 141 GDMPLAPLKEPLTPLISGLSGAMDQVTKMDRATGDLHTYISAPDLHMAASVYKHS 200
DB 377 G-----ANVTASTVSTSGT-TTVGAN--SAASG-----TVSVVANSS 412
QY 201 PSIAROVLDTGVAAVQT--YSARNAVRTVLAPALSRPAVQAVDGLVSMAGGLAANFG 258
DB 413 TTTTGAIAVTGTAIVTAQTRAGNAVNTTLTQADVTVTGNSSSTAIVTQTAAATGATVA 472
QY 259 NRLSVQRDHORGALVGLKDKPEPAQLSEENDMLEAVYKAISASVSGAALNAGRMA 318
DB 473 GRV-----NGAVTI-----TDAASAATTAGK--- 494
QY 319 GLPLDMATDMAGAV-RSLVSASSLTONGLALAGFAVGLQEMATKNITDPATKAASQ 377
DB 495 -----IITVILSGFGAATITSSALTIVNLSTGTSIGR-----GALT--ATPTANTL 541
QY 378 LTNLAGAAVPAWGTTALTTPPAVKAESFIQD---TVK---STASTTGVAVDQVYKL 431
DB 542 TLNANG-----LTTGATDSEAAADGFTTINAGTAST--IASLVAAD 586
QY 432 AKTVKDMGGEAIT--HTGASLRN-TVNN 456
DB 587 ATTINISGDARVTTTSHTAALTGITVTN 615

```

```

RESULT 4
US-08-591-079-6
; Sequence 6, Application US/08591079
; Patent No. 5972899
; GENERAL INFORMATION:
; APPLICANT: Zychlinksky, Arturo
; APPLICANT: Chen, Yajing

```

```

; TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FORSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,079
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Laynat, Shmuel
; REGISTRATION NUMBER: 33,949
; REFERENCE/DOCKET NUMBER: 15661-20017.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0764
; TELEX: 90-4030 MRSNFORSMH
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-591-079-6

```

```

Query Match      5.3%; Score 127.5; DB 2; Length 401;
Best Local Similarity 19.7%; Pred. No. 0.0012;
Matches 82; Conservative 57; Mismatches 124; Indels 153; Gaps 17;

```

```

QY 121 RATPFAMASLLQYMQ--PA-----INKGMLPAPLPLPLISGAL 159
DB 9 RSTP-VTGSLLPYETPPAPPLDTQOVAGSLKDKNGVSSGIVQLPAPL---AVASQV 63
QY 160 SGAMDQVTKMDRATGDL-----HYLS-----ASPD----- 186
DB 64 TEGQOQEVTLKLSVTRGAAGSQLISNVSLTKFTLASPDTEIEBKLVSNLEVRKD 123
QY 187 -----RLHDMAAASVYKHSPL-----ARQYLDTGVAVQTYSAANAVRTYLAAPLAS 233
DB 124 IKIADIORLHEONMKKIEBQEKIKETEENAKQVKSGISAKIFGWLISAISVYVGAIMV 183
QY 234 RPAVQAVDVGSMAGGLAANAGFGNRLISVQSRDHORGALVGLKDKPEPKAQLSEND 293
DB 184 ASGV-GAV-----AGAMVVASG-----YIGM----- 203
QY 294 WLEAYKAIKSASVGAALNAGKEMAGLPLDMATDMAGAVRSLVSASLTONGLALAGFA 353
DB 204 ---ANNAVQAAADGLISQAMKILGPILTALIEVALTVSVTWTFGSALKCLANIGAKL 260
QY 354 GVGKLQEMATKNITDPATKAASQ-LTNLAGSAAVPAWGTTALTTPPAVKAESFIQDT 412
DB 261 G-ANTASLAKGAEFSAKVQISTGISTNVGSAVTYKGGSGFAGITWSHAIR----- 310
QY 413 VKSTASTTGVAVDQVYKLAKTVKDMGGEAITHTGASLRNVTNNLRPAPAREADIE 468
DB 311 ---TGSQAT-----QVAVGVSGITOTINNKKO---ADLQ 339

```

```

RESULT 5
US-09-085-199B-7
; Sequence 7, Application US/09085199B
; Patent No. 6235879
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael R.

```

```

; APPLICANT: Hackam, Abigail
; APPLICANT: Huq, A.H.M. Mahbubul
; APPLICANT: Chopra, Vikramjit Singh
; APPLICANT: Kalchman, Michael
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the
; TITLE OF INVENTION: Huntington's Disease Gene
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Oppedahl & Larson
; STREET: PO Box 5270
; CITY: Frisco
; STATE: CO
; COUNTRY: USA
; ZIP: 80433-5270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS DOS 5.0
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,199B
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Larson, Marina T.
; REGISTRATION NUMBER: 32038
; REFERENCE/DOCKET NUMBER: UBC.P-013US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (970) 668-2050
; TELEFAX: (970) 668-2052
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 676
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: no
; ORIGINAL SOURCE:
; ORGANISM: human
; FEATURE:
; OTHER INFORMATION: Huntington-interacting protein
;
; US-09-085-199B-7

```

```

Query Match 5.3%; Score 127; DB 4; Length 676;
Best Local Similarity 21.4%; Pred. No. 0.003;
Matches 117; Conservative 68; Mismatches 209; Indels 154; Gaps 24;

Qy 33 SVSSDQREINAIADYLT--DHVFAAHLPPADSADGQAADVHNAQITAIETASRL 89
Db 144 SHTEOSKSELSSRLTTLAEKDLSGAVRQREADLLAAQSLVRETEA---ALSREQRSS 200
Qy 90 HFEGETPATIADTFKAERKL-DRLATTTSGALRATFFAMASILQYMQPAINKGWLPAFL 148
Db 201 QEQELQGLAERESQEQGLRQLLDEQFAVLRGAAEAAGIL---QDAVSKLD----- 251
Qy 149 KPL-----TP--LISGALSGAMDQVGYMMDRATGDLHYLSASPDRLHDMAASVKHS 200
Db 252 DPLHLRCTSSPDYLSRA-QEALDAVST---LEEGHAQYLTSLADA--SALVAALTRFS 304
Qy 201 PSLARQVLDTGAVQVTSYARNVAVTV-----LAPALAS--RPA 236
Db 305 HLAADIINGGATSHLPTDPADRLIDTCREGGARAELMGQLQDQQAALRMQASLVRTTP 364
Qy 237 VQGAVDLGVSMAGGLAANAGFGRNLLSVQSRDHQRGGALVLGLKDKPEKAQLSEENDWLE 296
Db 365 LQGIILQLOQL-----KPKSLDVRQEEIGAVV-----DKEMAATSAAIEDAVR 407
Qy 297 AYKAITKSA---SYSGAALNAGKRMAGLPLDMATDMGAVRSVLVSASSLTQGLALAGGFA 353
Db 408 RIEDMMNQARHASSGVKLEVNRI----LNSCTDLMKAIRLLVTTSTLQKEIVESG--R 461
Qy 354 GVGLQEMATKNITDPATKAAVSQLTNLGASAAVPAWGTTAALTTPDAVKAESPIQDVT 413

```

```

Db 462 GAATQOEFYAKN-----SRWTEGLISAKAVCWGATQL-----VEADKVVLTHT 506
Qy 414 K-----STASSTGYVADQTVK-----LAKTVKMGGBAITHT----- 446
Db 507 KYBELIVCSHEIAASTAQLVAASKVKANKHSPHLRLQECSTVNERRAANVASTKSGQE 566
Qy 447 -----GASLRNTVNNLRQRPAREADIEE-----GGTAASP 476
Db 567 QIEDRDMDFSGLSLIKKKOEMETQVRVLELEKLTLEAERMLGELRKQHYVLGASGSP 626
Qy 477 -SEIPRP 483
Db 627 GEEVAIRP 634

RESULT 6
US-08-968-685A-10
; Sequence 10, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: TUCKER, KENNETH
; APPLICANT: PLOSILA, LAURA
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,685A
; FILING DATE: No. 6214981ember 12, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine P.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2123 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-968-685A-10

```

```

Query Match 5.3%; Score 127; DB 4; Length 2123;
Best Local Similarity 21.1%; Pred. No. 0.019;
Matches 116; Conservative 84; Mismatches 218; Indels 132; Gaps 27;

Qy 23 DASLASSVRSVSSDQOREIN-----AIADYLTLDVHPAAHKLPPADS-----ADGQA 69
Db 281 DNDVNAANRAYAPDDNQPIDNRYKATFKNGATD-VFSIGNSGNDSIRRKIIINVAGSA 339
Qy 70 AVDVHN-AQITALIEETRSLRHEGETPATIADTFKAERK-LDRLATTTSCALRATPFAM 127
Db 340 DTDVAVNAQKLEAVRLANRQITFKGD-----DSNNRVEKGLGKTLTITGA-----QT 387
Qy 128 ASLIQYMQPAINKGDLPAFL-KPLTPL-----ISGALSGAMDQVQ----- 167
Db 388 SALTDHNIQVQNGDGLKVQLAETLTSLKMWTTENLTANKEVTVGKTRLTITDKIGFTNDM 447

```

QY 168 -----TKMDRATGDLHYLSASPRDLHDMAASVKHSPSLA--ROYLDTG-VAVQVYS 218
 Db 448 NGIDESKPYLDKOTG-----IHAGGQKITKLTAAGVDDDAATYGGOLKKVNGTAESALQFT 503
 QY 219 ARNAVRI-----VLAPALASRP-----AVQAVDLGVSMAGG 250
 Db 504 VKKYDKRGANDANDSKITTVGKNRPDGTQVNTLKLKGENGVDTTEGTTPFLPANNNG 563
 QY 251 L-AANAGFGRNLLSVQSRDHOR-----GGLVGLKDKPEKAOISENDLBAVKAIKS 303
 Db 564 LTVGNSTLNDGLSVKNTNSKKQIQVAGADGTTFPDINSKSGAGI--ENTRITRDTGIGF 621
 QY 304 ASYSGAALNKG-RMAGLPDMATDMAGVSLVSASSLTONGALAGFAGVGL--Q 359
 Db 622 ANNTG-SLDANKPRLTPTGINAGGKELTNVQSAINPAT--NG-----GQLDFMN 667
 QY 360 EMATKNTITDPAITKAASQULTLAGSAVFAQTALITDPAVKKASFIQDIYKSTRSS 419
 Db 668 RLSTANTEKSGSATITDLYNLGVPLTFAGDTGPNVTK-----KLGEILKVKGGKTTADD 723
 QY 420 TT-----GVVADQ-----TVKLAKTVKDMGGEATHTGTASLRNTV---NNLRORAREADI 467
 Db 724 LTKNIGIVVADSTNSLTIVLAKTILSDIDAVNTKTLTRASDKVTVDSGNNTAKLQNGDLTF 783
 QY 468 EEGSTAAPS 477
 Db 784 SKQNTGATPA 793
 RESULT 7
 US-09-071-035-264
 ; Sequence 264, Application US/09071035
 ; Patent No. 6448043
 ; GENERAL INFORMATION:
 ; APPLICANT: Gil H. Choi
 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 ; NUMBER OF SEQUENCES: 496
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/071,035
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: A. Anders Brookes
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB369P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 264:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 669 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-071-035-264

Query Match 5.2%; Score 126.5; DB 4; Length 669;
 Best Local Similarity 20.8%; Pred. No. 0.0033;
 Matches 113; Conservative 69; Mismatches 192; Indels 169; Gaps 27;
 QY 17 SFRT-ASDASLASSSVRSVSSDDOREINALADVLTHVFAHLPPRDSADGQAAVYHN 75
 Db 116 TFOFPDAEATVTAPTTIGTVG-----NSTAGY-----EVKGTADANAATYIRN 158
 QY 76 AQITALETRASRLHFECEPATIADTFAKA---EKDLRLATTTSGALRLATPMAISLQ 132
 Db 159 AG-GAVIGTGA-----DGTAFVTTIPAGAGANETITAVAKASG-ESTPTTFQI--- 209
 QY 133 YWQPAINKGDMPLAPLKLPL-----ISGALSGAM-----QVGT--RMMDRATGDL 178
 Db 210 -----PADPNTPVATPIVEVTGSTTKGYEVKGTAEVGTITIEVRDAAGTVL 255
 QY 179 HYLSASPRDLHDMAASVKHSPSLARQVLDTVAAV--QVYS--ARNAVTVLAPALASR 234
 Db 256 GTATGTGDKRYV-----TLDSGTATANQTLVAVAKASGSESQPATATT 300
 QY 235 PAVQAVDLGVSMAGGLAANAGFGRNL-----LSVQSRDHORGALVGLKDKPEKAQ 287
 Db 301 PA-----DVTAFTVDNITGSSGGEITGTADPNTTIEVRD-PSGAVIGTGTSD----- 348
 QY 288 LSEENDWLEAYKAIKSASYSGAALNAGKMAGLPDMATDMAGVSLVSA----- 338
 Db 349 --ANGDF-----TVTLPTGTTNPGDILLVIGKDNAGNESQTEVLVPADATVTAFTV 398
 QY 339 SSLTONGLAGFAGVAGKLOEMATKNTIDPATKAASQULTNLASNAV-----PAGW 391
 Db 399 TGVYGNSSVA---GYQVGTADPNATIEIRPADGNVIATGTADTGSFAVLLPAGTANANE 455
 QY 392 TTATLTDPVAKAESFIQDTV-----KSTASSTGVVADQVYVLAKTVK---- 426
 Db 456 TLRLAKDPAGNTSTPTTFOTPADEVVAPPSVDKVTGNTTQGTQVGTAEALGTTIEVRAT 515
 QY 437 ---DMGGEATHTG-----ASLRNTV-----NLORPARADIEEGSTAASP 476
 Db 516 DGTVLGTATGPGQYTVTLASGKATAKQVNVVAVKNDTGLBSQPT-----TAMTP 566
 QY 477 SEI 479
 Db 567 ADV 569
 RESULT 8
 US-08-614-377A-7
 ; Sequence 7, Application US/08614377A
 ; Patent No. 5976864
 ; GENERAL INFORMATION:
 ; APPLICANT: Smit, John
 ; APPLICANT: Bingle, Wade H.
 ; TITLE OF INVENTION: EXPRESSION AND SECRETION OF
 ; TITLE OF INVENTION: HETEROLOGOUS
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson PC
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/614,377A
 ; FILING DATE: 12-MAR-1996
 ; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/194,290
 ; FILING DATE: 09-FEB-1994
 ; CLASSIFICATION: 435
 ; APPLICATION NUMBER: US 07/895,367
 ; FILING DATE: 09-JUNE-1992
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Tsao, Y. Rocky
 ; REGISTRATION NUMBER: 34053
 ; REFERENCE/DOCKET NUMBER: 08106/002001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-542-5070
 ; TELEFAX: 617-542-8906
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1026 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-614-377A-7

Query Match 5.2%; Score 126.5; DB 2; Length 1026;
 Best Local Similarity 22.2%; Pred. No. 0.0066;
 Matches 113; Conservative 56; Mismatches 177; Indels 163; Gaps 23;

 QY 12 VTATDSPTASDASLASSSVRSVSSDQOREINAIADY----- 48
 DB 206 VSGIGGYATATAAMINDLSGALSTDNAAGVNLFTAYPSSGVSGSTLSLTGTDLTGTA 265

 QY 49 -----LTDHVFAAHKLPADSDAGQAAVDVHN-----AQITAL-----IETRASRLHF 91
 DB 266 NNDTEVAGEVAGAAATLVGDTLGGAGTDLNWNQAAAVTALPTGVTISGIET-----MNV 321

 QY 92 EGEPATTAIDTFAKAEKLDRLATTTSQA-----LRATPFAMASLLQYMPAINK 140
 DB 322 TSGAAITL-NTSSGVGTGLTALNTNTSGAAQVTAGAGQNLTAATAQAANNVVDGRAN- 379

 QY 141 GDWLPAPLKPPLTISGALSGAMDQVTKMDRATGDLHYLSASPDRLHDAMAASVKRHS 200
 DB 380 -----VTVAAGTGTSGT-TTVGAN--SAASG-----TVSVSVANSS 412

 QY 201 PSLARQVLDTGAVQT--YSARNVAVTLAPALASRPVQGVAVDGLVSMAGGLAANAGFG 258
 DB 413 TTTTGAIAVTGGTAVTVAQTAGNAVNTLTQADVTTGNSSTTAVTQTAAATAGATVA 472

 QY 259 NRLLSVQSRDHRQAGALVGLKDKPEKQALSEENDWLEAYKAIKASYSYSGAALNAGKMA 318
 DB 473 GRV-----NGAVTI-----TDSAAASATTAGK--- 494

 QY 319 GLPLDMATDMGAV-RSLVSGASSLTONGLAGGAGVGLQEMATKNIITDPATKAAVSQ 377
 DB 495 -----IATVTLGSFGAATIDSSALTTVNLSTGTSGLGR-----GALT--ATPTANTL 541

 QY 378 LTNLAGSAAVFAGWTTAALTTPDAVKAESFIQD---TVK---STASSTTGYVADQTVKL 431
 DB 542 TLNVNG-----LTTGATDSEAAADGFTTINAGSTASST---IASLVAAD 586

 QY 432 AKTVKDMGGEAIT---HTGASLRN-TVNN 456
 DB 587 ATTILNISGDARVTITSHATAALTGITVTN 615

RESULT 9
 US-09-142-648B-7
 ; Sequence 7, Application US/09142648B
 ; Patent No. 6210948
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, John
 ; APPLICANT: Bingle, Wade H.
 ; APPLICANT: No. 6210948ellini, John F.
 ; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULO

; FILE REFERENCE: 08106/002002
 ; CURRENT APPLICATION NUMBER: US/09/142,648B
 ; CURRENT FILING DATE: 1999-03-30
 ; PRIOR APPLICATION NUMBER: PCT/CA97/00167
 ; PRIOR FILING DATE: 1997-03-10
 ; PRIOR APPLICATION NUMBER: US 07/614,377
 ; PRIOR FILING DATE: 1996-03-12
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 1026
 ; TYPE: PRT
 ; ORGANISM: Caulobacter crescentus
 ; US-09-142-648B-7

 Query Match 5.2%; Score 126.5; DB 4; Length 1026;
 Best Local Similarity 22.2%; Pred. No. 0.0066;
 Matches 113; Conservative 56; Mismatches 177; Indels 163; Gaps 23;

 QY 12 VTATDSPTASDASLASSSVRSVSSDQOREINAIADY----- 48
 DB 206 VSGIGGYATATAAMINDLSGALSTDNAAGVNLFTAYPSSGVSGSTLSLTGTDLTGTA 265

 QY 49 -----LTDHVFAAHKLPADSDAGQAAVDVHN-----AQITAL-----IETRASRLHF 91
 DB 266 NNDTEVAGEVAGAAATLVGDTLGGAGTDLNWNQAAAVTALPTGVTISGIET-----MNV 321

 QY 92 EGEPATTAIDTFAKAEKLDRLATTTSQA-----LRATPFAMASLLQYMPAINK 140
 DB 322 TSGAAITL-NTSSGVGTGLTALNTNTSGAAQVTAGAGQNLTAATAQAANNVVDGRAN- 379

 QY 141 GDWLPAPLKPPLTISGALSGAMDQVTKMDRATGDLHYLSASPDRLHDAMAASVKRHS 200
 DB 380 -----VTVAAGTGTSGT-TTVGAN--SAASG-----TVSVSVANSS 412

 QY 201 PSLARQVLDTGAVQT--YSARNVAVTLAPALASRPVQGVAVDGLVSMAGGLAANAGFG 258
 DB 413 TTTTGAIAVTGGTAVTVAQTAGNAVNTLTQADVTTGNSSTTAVTQTAAATAGATVA 472

 QY 259 NRLLSVQSRDHRQAGALVGLKDKPEKQALSEENDWLEAYKAIKASYSYSGAALNAGKMA 318
 DB 473 GRV-----NGAVTI-----TDSAAASATTAGK--- 494

 QY 319 GLPLDMATDMGAV-RSLVSGASSLTONGLAGGAGVGLQEMATKNIITDPATKAAVSQ 377
 DB 495 -----IATVTLGSFGAATIDSSALTTVNLSTGTSGLGR-----GALT--ATPTANTL 541

 QY 378 LTNLAGSAAVFAGWTTAALTTPDAVKAESFIQD---TVK---STASSTTGYVADQTVKL 431
 DB 542 TLNVNG-----LTTGATDSEAAADGFTTINAGSTASST---IASLVAAD 586

 QY 432 AKTVKDMGGEAIT---HTGASLRN-TVNN 456
 DB 587 ATTILNISGDARVTITSHATAALTGITVTN 615

RESULT 10
 US-09-071-035-258
 ; Sequence 258, Application US/09071035
 ; Patent No. 6448043
 ; GENERAL INFORMATION:
 ; APPLICANT: Gil H. Choi
 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 ; NUMBER OF SEQUENCES: 496
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage


```

1 COMPUTER: HP Vectra 486/33
2 OPERATING SYSTM: MSDOS version 6
3 SOFTWARE: ASCII Text
4
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/09/071,035
7 FILING DATE:
8 CLASSIFICATION:
9
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER:
12 FILING DATE:
13 ATTORNEY/AGENT INFORMATION:
14 NAME: A. Anders Brookes
15 REGISTRATION NUMBER: 36,373
16 REFERENCE/DOCKET NUMBER: PB369P2
17
18 TELECOMMUNICATION INFORMATION:
19 TELEPHONE: (301) 309-8504
20 TELEFAX: (301) 309-8512
21 INFORMATION FOR SEQ ID NO: 258:
22
23 SEQUENCE CHARACTERISTICS:
24
25 LENGTH: 1638 amino acids
26 TYPE: amino acid
27 STRANDEDNESS: single
28 TOPOLOGY: linear
29
30 MOLECULE TYPE: protein
31
32 US-09-071-035-258

```

Query Match 5.2%; Score 126.5; DB 4; Length 1638;

```

QY 17 SPRT-ASDASLASSSVRSVSDQDEINADYLDTHVPAHKLPPADSDAQOAAVDHN 75
Db 523 TFQTPADEYVAPPTTGVGT---NSTAGY-----EYKGDANAVATEIEN 565
QY 76 AQITALLIETRASRLHFEGEPTPATIADTFEKA--EKUDRLATTTSGALRATPEFAMASLIQ 132
Db 566 AG-GAVIGTGA---DGIQAFVTIIPAGEAGANEITLTVAKNASGT-ESTPTTPTQ--- 616
QY 133 YMOPAINKGDWLPAPLKLPTPL-----ISGALSGAMD---QYGT--KMDRATGDL 178
Db 617 -----PADPPTPVATPPIVETVTSTTKGYEVKGTAEVGTIIEVRDAGTVL 662
QY 179 HYLASPRRLHDAMAAVSRHSPLBARQVLDYGVAV--QTVS--AENAVTYLAPLASR 234
Db 663 GTATGTGDKTYV-----TLDSGTANQTLISVNAKNAAGTSSQPTATT 707
QY 235 PAVQGAVDLGVSMAGGLANAGFGRNL-----LSVQSRDHORGALVIGLKDKEPKAQ 287
Db 708 PA-----DVYAPTVNITGNSSGYEITGTADPNTTIEVRD-PSGAVIGTGSID----- 755
QY 288 LSEENDWLEAVYKAISASYSGAALNAGKEMAGIPLDMATDANGAVRSLSVA----- 338
Db 756 --ANGDF-----TVTLPTGTNPBGDTLTVIGKDNAGNESQPEVLVPADATVYAPTV 805
QY 339 SSLTONGLALAGFGAGVGLQEMATKNITDPATPKAAVSQLTNLASGAV-----FAWM 391
Db 806 TGVNTNSVA--GYVVTGTADPNATITEIRDADGNVYATATAGTGSFAVNLPAGTANANE 862
QY 392 TTAALTTPPAVKKASEFIQDTV-----KSTASSTGVADQVTKLAKTVK----- 436
Db 863 TLTALAKCPAGNTSPPTTPTQTPADEVVAAPSPVDKGTGNTTQGYQVGTLELGTITIVRAT 922
QY 437 ---DMGGEALITHYG-----ASLRNTVN-----NLQRPARREADIEGGTAAAP 476
Db 923 DGTVLGTATTTGTYVTVTLASGAKATQKTVNVVAKNDIGLSSQPF-----TAMRP 973
QY 477 SEI 479
Db 974 ADV 976

```

RESULT 11
US-09-071-035-262

```

Sequence 262, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
CLASSIFICATION:
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ. ID NO: 262:
SEQUENCE CHARACTERISTICS:
LENGTH: 1638 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-071-035-262

```

Query Match 5.2%; Score 126.5; DB 4; Length 1638;
 Best Local Similarity 20.8%; Pred. No. 0.014;
 Matches 113; Conservative 69; Mismatches 192; Indels 169; Gaps 27;

QY	17	SFRF--ASDASLASSSVSSVSDQOREINALADVLTHVFAAHKLPRADSPADGQAAVDVHN	75
Db	523	TFQFPDPAEATVATPTTGVG-----NSTAGY-----EYKGDANATWEIRN	5655
QY	76	AQIALLIETRASRLHFEGEPATIIDPFAA-----EKIDRLATTSGLRATPFAMSLQ	1322
Db	566	AG-GAVIGTGTA-----DGTGAFYTTIPAGSAGANETLTVAANKASGT-ESTPTTTFQF--	6166
QY	133	YMQPAINKGDMPLPKPLPLPL-----ISGALSGAMD-----QVGT--KXMDRATGDL	1786
Db	617	-----PADPNTPFVATPIVEITVGTSTTKCYEKVGTAEVGTITTEVDAAGTVL	6623
QY	179	HYLSASPRDLHDMAAASVKKHSPSLANQVLDTVAAV--QTVS--ARNAARVTLAPLALSR	2354
Db	663	GTATTTGTGDKTV-----TLDSGFATRNQTLISVAAKASGTESQOPATATT	7077
QY	235	PAVOGAVDLGVSMAGLAAANAFGNRL-----LSVQRSDHORGALVLGLKDKEPKAQ	2873
Db	708	PA-----DYTAFTPTDNTIGNSGSEYELTGTADPNTTIEVAD-PSGANVIGTGTSD-	7555
QY	288	LSEENDMLEAVKAIKASYSYGALINACKRMAGLP.LDMATDAMGAVNSLYSA-----	3388
Db	756	--ANGDF-----ITVLTPTGTTNPPEDTLTVIGCKNAGNESOPTREVLVPADATVTAFTV	8055
QY	339	SSLQNGGLALAGSGAVGKQKQEMATKNITTPATKAASVQLTINLAGSAV-----PAGW	391
Db	806	TGVIGNSVVA-----GQVTTGTADPNNAETIEIRADGQNVATGTADGTGSAVNLPACTANANE	8623

QY 392 TTAALTTPAVKKAESFIQDITV-----KSTASSTTGYVADQTVKLAKTVK----- 436
DB 863 TLTALAKDPAGNTSTPTTTPADDEVVAPPVSKVTGNTTQGYQVGTGAELGTTIEVRAT 922
QY 437 ---DMGGEAITHTG-----ASLRNTVN-----NLQRPAAREADIEEGGTAASP 476
DB 923 DGTVLGTATTGPTGYTVTLASGRATKATQTVNVVAKNDTGLSQPT-----TAMTP 973
QY 477 SEI 479
DB 974 ADV 976
RESULT 12
US-09-071-035-266
; Sequence 266, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 266:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-266
Query Match 5.2%; Score 126.5; DB 4; Length 1638;
Best Local Similarity 20.8%; Pred. No. 0.014;
Matches 113; Conservative 69; Mismatches 192; Indels 169; Gaps 27;
QY 17 SFRT-ASDASLASSSVRSVSSDQOEINAIADYLTDFVFAHKLPADSDAQAAVDVHN 75
DB 523 TFQTPADEATVAPTITGTG-----NSTAGY-----EVKGTADANATVEIRN 565
QY 76 AQITALIESTRASLHFEGETPATIADTFKA---EKLRLATTISGALRAITPFAMASLLQ 132
DB 566 AG-GAVIGTGA-----DGTGFTVTIPAGEAGANTLTAVAKASGT-ESTFTTFTQ--- 616
QY 133 YNQPAINKGDMPLAPLKLPTPL-----ISGALSGAMD-----QVGT--KMDRATGDL 178
DB 617 -----PADPNTPVATPIVETVGTSTTKGVEVKGTAEVGTITIEVRDAAGTVL 662
QY 179 HYLSPDRHLHDAASVKRHSPLSLARQLDTGVAV--QTYS--ARNAVRTVLAPALASR 234

DB 663 GTATTGTGKTV-----TLDSGTATANQTLNVVAKNAGSTESQPATAIT 707
QY 235 PAVQGVADVLGVSVMAGGLAANAGFNRL-----LSVQSRDHQHGALVLGLKDKKEPKAQ 287
DB 708 PA-----DVTAPTVDNITGNSGSGYEITGTADPNTTIEVRD-PSGAVIGTGTSD----- 755
QY 288 LSEENDWLEAYKAIKASYSYGAALNAGKRMAGLPLDMATDAMGAVRSLVSA----- 338
DB 756 --ANGDF-----TVTLPTGTTFNFGDTLTVIGKONAGNESQFTEVLVPADATVTAPT 805
QY 339 SSLTQGLALAGGAGVGKLEQMATKNIITDPATKAAVSQLTNLGSAAV-----PAGW 391
DB 806 TGVIGNSVA---GYQTGTADPNATIEIRDADGNVIAITGADGTGSAFVNLPGATNANE 862
QY 392 TTAALTTPAVKKAESFIQDITV-----KSTASSTTGYVADQTVKLAKTVK----- 436
DB 863 TLTALAKDPAGNTSTPTTTPADDEVVAPPVSKVTGNTTQGYQVGTGAELGTTIEVRAT 922
QY 437 ---DMGGEAITHTG-----ASLRNTVN-----NLQRPAAREADIEEGGTAASP 476
DB 923 DGTVLGTATTGPTGYTVTLASGRATKATQTVNVVAKNDTGLSQPT-----TAMTP 973
QY 477 SEI 479
DB 974 ADV 976
RESULT 13
US-09-085-199B-11
; Sequence 11, Application US/09085199B
; Patent No. 6235879
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael R.
; APPLICANT: Hackam, Abigail
; APPLICANT: Hug, A. H. M. Mahbubul
; APPLICANT: Chopra, Vikramjit Singh
; APPLICANT: Kaichman, Michael
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the
; TITLE OF INVENTION: Huntington's Disease Gene
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppedahl & Larson
; STREET: PO Box 5270
; CITY: Frisco
; STATE: CO
; COUNTRY: USA
; ZIP: 80443-5270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS DOS 5.0
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,199B
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Larson, Marina T.
; REGISTRATION NUMBER: 32038
; REFERENCE/DOCKET NUMBER: UBC.P-013US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (970) 668-2050
; TELEFAX: (970) 668-2052
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: no
; ORIGINAL SOURCE:
; ORGANISM: mouse

Job time : 20.4861 secs

```
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: JP 224711/94
  FILING DATE: 20-SEP-1994
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: JP 106006/95
  FILING DATE: 28-APR-1995
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: JP 106008/95
  FILING DATE: 28-APR-1995
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: JP 106009/95
  FILING DATE: 28-APR-1995
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: JP 106010/95
  FILING DATE: 28-APR-1995
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: JP 106011/95
  FILING DATE: 28-APR-1995
ATTORNEY/AGENT INFORMATION:
  NAME: Miller, Charles E.
  REGISTRATION NUMBER: 24,576
  REFERENCE/DOCKET NUMBER: 7426-043-999
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (212) 790-9090
  TELEFAX: (212) 869-8864/9741
  TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
  LENGTH: 488 amino acids
  TYPE: amino acid
  TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-809-326A-1

Query Match      5.1%; Score 122; DB 4; Length 488;
Best Local Similarity 21.2%; Pred. No. 0.0053;
Matches 102; Conservative 62; Mismatches 168; Indels 148; Gaps 22;

Qy 25 SLASSSVRSVSDQREINAIADYLTDRHFAAHLPPADSDADGQAADVHVNAQTALLET 84
Db 2 SISSSS---GPDNQK--NIMSVLTS---TPQGVPOQDKLSGNETKQIQ-----QT 44

Qy 85 RASRLHFEGETPATIADTFKAKEKLDRLATTT----- 116
Db 45 RQGR-NTEMESDATTAGSGK----DKTSSTTKTETAPQQGVAAAGESSESQKAGADTV 99

Qy 117 SGALRATPFAMASLIQVMPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTQMDRATG 176
Db 100 SGAAATTSATATKTA-NQTSIEE-----ASKSMESTLE 132

Qy 177 DLHYLSASPDRLHDAM-----AASVKRHSPSLAR-----QVLDTGAVAVQVYSAR 220
Db 133 SLQSLSAQAQMEVEAVVAALSGKSGSAKLETPELPKPGVTPRSEVIEIGLAL-----A 187

Qy 221 NAVRTV---LAPALASRAVQGVAVDLGVSMAGGLAANAGFGRNLLSVQSRDHQRCALVL 277
Db 188 KAIQTLGRATKSALSNSYASTQAQD-----QTNKLGLEKQAIKIDKEREYQEMKAA 239

Qy 278 GKQKPEKQALSEENDWLEAYK-AIKSASYSGAALNAGKRMAGLPDMDATDAMGAVRSLV 336
Db 240 EOKSKDLEGTMDTVNTVMIASVAITVISIVAAIFTCAGLAGL-----AAGNA---- 288

Qy 337 SASSITQNLALAGFAGVGKLOEMATKNIITDPATKAAVSQITNLAGSAAVFAAGTTTAAAL 396
Db 289 -----VGAAAAGAGAGAAATTTATQ-IITQAVVQAVKQAVITAVRQAI-----TAAI 335

Qy 397 TTDPAVKA-ESEFQDTVKSTASST-----TCGVADQTVKLAKTVKDMGGEAIT 444
Db 336 KA--AVKSGIKAFIKTLVKAIAKAIKSGISKVFAKGTQMIKPNPKLSKVISLTSKWT 393
```

Search completed: January 2, 2003, 15:21:41

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 2, 2003, 15:13:55 ; Search time 34.5355 Seconds
(without alignments)
1879.027 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407
Sequence: 1 RHINOSAQQPGVAMESEFRT.....EEGGISAFSRETPQLRRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seque, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID22/gcgdata/geneeq/geneeq-emb1/AA1980.DAT:*
2: /SID22/gcgdata/geneeq/geneeq-emb1/AA1981.DAT:*
3: /SID22/gcgdata/geneeq/geneeq-emb1/AA1982.DAT:*
4: /SID22/gcgdata/geneeq/geneeq-emb1/AA1983.DAT:*
5: /SID22/gcgdata/geneeq/geneeq-emb1/AA1984.DAT:*
6: /SID22/gcgdata/geneeq/geneeq-emb1/AA1985.DAT:*
7: /SID22/gcgdata/geneeq/geneeq-emb1/AA1986.DAT:*
8: /SID22/gcgdata/geneeq/geneeq-emb1/AA1987.DAT:*
9: /SID22/gcgdata/geneeq/geneeq-emb1/AA1988.DAT:*
10: /SID22/gcgdata/geneeq/geneeq-emb1/AA1989.DAT:*
11: /SID22/gcgdata/geneeq/geneeq-emb1/AA1990.DAT:*
12: /SID22/gcgdata/geneeq/geneeq-emb1/AA1991.DAT:*
13: /SID22/gcgdata/geneeq/geneeq-emb1/AA1992.DAT:*
14: /SID22/gcgdata/geneeq/geneeq-emb1/AA1993.DAT:*
15: /SID22/gcgdata/geneeq/geneeq-emb1/AA1994.DAT:*
16: /SID22/gcgdata/geneeq/geneeq-emb1/AA1995.DAT:*
17: /SID22/gcgdata/geneeq/geneeq-emb1/AA1996.DAT:*
18: /SID22/gcgdata/geneeq/geneeq-emb1/AA1997.DAT:*
19: /SID22/gcgdata/geneeq/geneeq-emb1/AA1998.DAT:*
20: /SID22/gcgdata/geneeq/geneeq-emb1/AA1999.DAT:*
21: /SID22/gcgdata/geneeq/geneeq-emb1/AA2000.DAT:*
22: /SID22/gcgdata/geneeq/geneeq-emb1/AA2001.DAT:*
23: /SID22/gcgdata/geneeq/geneeq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2407	100.0	487	22	AAE12603
2	1869.5	77.7	486	22	AAE12573
3	1869.5	77.7	486	22	AAE12573
4	147	6.1	1463	23	AAE20110
5	146.5	6.1	2123	22	AAE00701
6	140	5.8	2285	20	AAW98149
7	138	5.7	1229	22	ABR17136
8	138	5.7	1277	22	AAE76536
9	138	5.7	2012	22	AAE76534
10	138	5.7	2993	22	AAE92485

11	136	5.7	1026	15	AAE48993
12	132	5.5	1026	18	AAW37490
13	132	5.5	1026	21	AAW44757
14	130.5	5.4	2368	22	AAU34139
15	130.5	5.4	2368	22	AAU36796
16	130	5.4	1026	17	AAE94014
17	129.5	5.4	2076	22	AAU34319
18	129.5	5.4	2186	22	AAU37320
19	129	5.4	2444	22	AAU37120
20	129	5.4	6025	22	AAE84939
21	127.5	5.3	2387	22	AAU01183
22	127.5	5.3	2478	22	AAU34320
23	127.5	5.3	2478	22	AAU37374
24	126	5.2	727	22	ABE25530
25	125.5	5.2	623	23	AAU12038
26	125.5	5.2	1252	16	AAE80530
27	125	5.2	456	22	ABE64582
28	125	5.2	793	22	AAU27571
29	125	5.2	793	22	AAE10016
30	125	5.2	2086	22	AAU34143
31	125	5.2	5795	22	AAU37017
32	124	5.2	628	23	AAE66009
33	124	5.2	3241	23	AAE66005
34	124	5.2	3647	11	AAE05041
35	123.5	5.1	2434	22	AAU34339
36	123.5	5.1	6281	22	AAU37403
37	123	5.1	230	22	ABE58344
38	123	5.1	688	22	ABE58317
39	123	5.1	232	22	AAV57274
40	123	5.1	1328	20	AAV20088
41	122.5	5.1	1040	23	ABE40520
42	121.5	5.0	1068	21	AAV59273
43	121.5	5.0	1787	23	ABE49791
44	120.5	5.0	548	22	ABE62090
45	120.5	5.0	810	22	AAE81232

ALIGNMENTS

RESULT 1	AAE12603	standard; Protein; 487 AA.
ID	AAE12603	
AC	AAE12603	
XX		
DT	03-JUN-2002	(first entry)
XX		
DE	Pseudomonas syringae pv. tomato strain DC3000 HopPro2 protein.	
XX		
KW	Conserved Effector Loc1; CEL; cytosolic; antibacterial; gene therapy;	
KW	Exchangeable Effector Loc1; EEL; disease resistance; transgenic plant;	
KW	eukaryotic cell death; cancer.	
XX		
OS	Pseudomonas syringae.	
XX		
PN	MO200175066-A2.	
XX		
PD	11-OCT-2001.	
XX		
PF	03-APR-2001; 2001WO-US10698.	
XX		
PR	03-APR-2000; 2000US-194160P.	
PR	11-AUG-2000; 2000US-324604P.	
PR	17-NOV-2000; 2000US-249548P.	
PA	(CORR) CORNELL RES FOUND INC.	
PA	(UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.	
PA	(UYNE-) UNIV NEBRASKA.	
XX		
PI	Collmer A, Alfano JR, Charkowski NO;	
XX		
DR	WPI; 2001-639361/73.	

resA-S-lyser prote
Caulobacter cresce
Caulobacter cresce
Staphylococcus aur
Staphylococcus aur
Caulobacter S-lyse
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Shrimp white spot
Rat glutamate tran
Staphylococcus aur
Staphylococcus aur
Novel human diagno
Clostridium diffic
B. sphaericus SLP.
Drosophila melanog
Neisseria meningit
N. meningitidis st
Staphylococcus aur
Staphylococcus aur
F. necrophorum tru
F. necrophorum leu
Filamentous haemag
Staphylococcus aur
Staphylococcus aur
Drosophila melanog
Drosophila melanog
E. carls 120 kda i
B. burgdorferi ant
Staphylococcus epi
Mouse huntingtin-1
Listeria monocycog
Drosophila melanog
Mycobacterium tube

Sam

DR N-PSDB; AAD20438.
 XX New nucleic acid molecules encoding proteins or polypeptides of
 PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
 PT genomic sequences, for imparting disease resistance to plants -
 XX
 XX Claim 8; Page 65-66; 217pp; English.
 XX
 XX The invention relates to an isolated nucleic acid molecule comprising a
 CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
 CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
 CC genomic sequences. CEL and EEL DNA are useful for imparting disease
 CC resistance to a plant, by transforming a plant cell with the nucleic acid
 CC and regenerating a transgenic plant from the transformed plant cell,
 CC where the transgenic plant expresses a heterologous DNA molecule under
 CC conditions effective to impart disease resistance, or by treating a plant
 CC with an isolated protein or polypeptide, by applying the protein or
 CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
 CC which secretes the protein or polypeptide, under conditions effective to
 CC impart disease resistance to the treated plant. CEL and EEL proteins
 CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
 CC Pseudomonas protein into a eukaryotic cell under conditions effective to
 CC cause cell death. CEL and EEL proteins are also useful for treating a
 CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
 CC cancer cells of a patient under conditions effective to cause death of
 CC cancer cells, and thus treating the cancerous condition. The method
 CC further involves administering a targeted DNA delivery system
 CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
 CC to the patient, where the targeted DNA delivery system delivers the
 CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
 CC is expressed in the cancer cells. The present sequence is
 CC Pseudomonas syringae pv. syringae HopPSYA homolog protein.
 XX
 SQ Sequence 487 AA;
 Query Match 100.0%; Score 2407; DB 22; Length 487;
 Best Local Similarity 100.0%; Pred. No. 9.4e-192;
 Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHINSAOQPPGVAMESFRTASDASLASSSVRSVSTTSCRDQLAITDYLKHHVFAAHRFS 60
 DB 1 MHINSAOQPPGVAMESFRTASDASLASSSVRSVSTTSCRDQLAITDYLKHHVFAAHRFS 60
 QY 61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
 DB 61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
 QY 121 ENTFFAASVLYQMOPAINKGDWLTPLKPLTSLGSLGAMDOVGTMMDRAGDILHY 180
 DB 121 ENTFFAASVLYQMOPAINKGDWLTPLKPLTSLGSLGAMDOVGTMMDRAGDILHY 180
 QY 181 LSTSPDKLHDAMAVSVKHSHPALGRQVVDGMIATVTFALNVVRLPALASRPSVOGA 240
 DB 181 LSTSPDKLHDAMAVSVKHSHPALGRQVVDGMIATVTFALNVVRLPALASRPSVOGA 240
 QY 241 VDFGVSTAGGLVANAGFGDRLSVQSRDQLRGAFVLGKMKKEPKAALSEETDMLDAYKA 300
 DB 241 VDFGVSTAGGLVANAGFGDRLSVQSRDQLRGAFVLGKMKKEPKAALSEETDMLDAYKA 300
 QY 301 IKSASYSGAALNAGMAGLPLDVATDGLKAVRSLSVATSITKNGLALAGGVAGVSKLOK 360
 DB 301 IKSASYSGAALNAGMAGLPLDVATDGLKAVRSLSVATSITKNGLALAGGVAGVSKLOK 360
 QY 361 MATKNITDSATKAASQLSNLVGSVGFAGWTAGTAGTDPVAKKAESFIQDKVKSTASST 420
 DB 361 MATKNITDSATKAASQLSNLVGSVGFAGWTAGTAGTDPVAKKAESFIQDKVKSTASST 420
 QY 421 TSYVADQTVKLAKTVKMSGEAISSTGASLSTVNNLRHSRAPEADIEGGISAPSRSET 480
 DB 421 TSYVADQTVKLAKTVKMSGEAISSTGASLSTVNNLRHSRAPEADIEGGISAPSRSET 480
 QY 481 PFQLRRL 487
 |||||

DB 481 PFQLRRL 487
 RESULT 2
 AAE12573
 ID AAE12573 standard; Protein; 486 AA.
 XX
 AC AAE12573;
 XX
 DT 03-JAN-2002 (first entry)
 XX
 DE Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 encoded protein.
 XX
 KW Conserved Effector Loci; CEL; cytostatic; antibacterial; gene therapy;
 KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
 KW eukaryotic cell death; cancer.
 XX
 OS Pseudomonas syringae.
 XX
 PN WO200175066-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 03-APR-2001; 2001WO-US10698.
 XX
 PR 03-APR-2000; 2000US-194160P.
 PR 11-AUG-2000; 2000US-224604P.
 PR 17-NOV-2000; 2000US-249548P.
 XX
 PA (CORR) CORNELL RES FOUND INC.
 PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
 PA (UYNE-) UNIV NEBRASKA.
 XX
 PI Collmer A, Alfano JR, Charkowski AO;
 XX
 WI: 2001-639361/73.
 DR N-PSDB; AAD20408.
 XX
 PT New nucleic acid molecules encoding proteins or polypeptides of
 PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
 PT genomic sequences, for imparting disease resistance to plants -
 XX
 PS Claim 8; Page 21-23; 217pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule comprising a
 CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
 CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
 CC genomic sequences. CEL and EEL DNA are useful for imparting disease
 CC resistance to a plant, by transforming a plant cell with the nucleic acid
 CC and regenerating a transgenic plant from the transformed plant cell,
 CC where the transgenic plant expresses a heterologous DNA molecule under
 CC conditions effective to impart disease resistance, or by treating a plant
 CC with an isolated protein or polypeptide, by applying the protein or
 CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
 CC which secretes the protein or polypeptide, under conditions effective to
 CC impart disease resistance to the treated plant. CEL and EEL proteins
 CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
 CC Pseudomonas protein into a eukaryotic cell under conditions effective to
 CC cause cell death. CEL and EEL proteins are also useful for treating a
 CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
 CC cancer cells of a patient under conditions effective to cause death of
 CC cancer cells, and thus treating the cancerous condition. The method
 CC further involves administering a targeted DNA delivery system
 CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
 CC to the patient, where the targeted DNA delivery system delivers the
 CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
 CC is expressed in the cancer cells. The present sequence is
 CC Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 protein.
 XX
 SQ Sequence 486 AA;
 Query Match 77.7%; Score 1869.5; DB 22; Length 486;
 Best Local Similarity 78.4%; Pred. No. 5e-147;

fermentation process; anti-infection; rotavirus infection; heart disease; infantile diarrhoea; lactose digestion; anti-cancer; autoimmune disorder; anti-mutagenesis; immune system modulation; allergy; Helicobacter pylori; antihypertensive effect; urogenital infection; hepatic encephalopathy; bowel syndrome; endocarditis; transgenic microbe; outer membrane protein; rompa.

Lactobacillus rhamnosus HN001.

WO200212506-A1.

14-FEB-2002.

08-AUG-2001; 2001WO-NZ00160.

08-AUG-2000; 2000US-0634238.

28-NOV-2000; 2000US-0724623.

(GENE-) GENESIS RES & DEV CORP LTD.

(VIAL-) VIALACTIA BIOSCIENCE NZ LTD.

Glenn M, Havukkala IJ, Bloksberg LN, Lubbers MW, Dekker J;

Christensson AC, Holland R, O'Loole PW, Reid JR, Coolbear T;

WPI; 2002-241760/29.

N-PSDB; AAD31881.

New polynucleotides and polypeptides from Lactobacillus rhamnosus, useful in e.g. improving the flavor, aroma, texture and health-related benefits of milk-derived products, or in increasing properties of microbes -

Claim 2; Fig 64; 257pp; English.

The present invention relates to a new isolated polynucleotide comprising a sequence present in Lactobacillus rhamnosus strain HN001 and encoding a polypeptide capable of modifying the flavour, aroma, texture, nutritional and health benefits of milk-derived products, and/or survivability of microbes in dairy manufacturing processes. The polynucleotides are useful for improving the properties of microbes used in the manufacture of milk-derived products such as cheeses, yogurt, fermented milk products, sour milks and buttermilk; in modifying the flavour, aroma, texture and health -related benefits of milk-derived products and in increasing the survival of microbes during industrial fermentation processes. The bacteria may be used to increase resistance to enteric pathogens and anti-infection activity, including treatment of rotavirus infection and infantile diarrhoea; aid in lactose digestion; as anti-cancer and anti-mutagenesis; liver cancer reduction; reduction of small bowel bacterial overgrowth; immune system modulation and treatment of autoimmune disorders and allergies; treatment of allergic responses to foods; reduction of blood lipids and prevention of heart disease; antihypertensive effect; prevention and treatment of urogenital infections, Helicobacter pylori, or hepatic encephalopathy; treatment of inflammatory bowel disorder and irritable bowel syndrome; modulation of endocarditis; and for improved protein and carbohydrate utilization and conversion. The transgenic microbial population can be administered to a mammal as an anti-carcinogenic agent. The present sequence is Lactobacillus rhamnosus outer membrane protein rompa.

Sequence 1463 AA;

Query Match 6.1%; Score 147; DB 23; Length 1463;

Best Local Similarity 20.8%; Pred. No. 0.005;

Matches 104; Conservative 64; Mismatches 241; Indels 92; Gaps 12;

Qy 1 MHINQSQPPGVAMESFRTASDASLASSVRSVSTTSRDIQAITDYLKHVFAHRFS 60

Db 714 INANSAADVTAQAKATSAASDAASYASEAQSIAGSHADNME-----IKSLASDAEQS 768

Qy 61 VIGSPDERDAALAHNEQIDALVETRNRLYSEGETPATI-----AETFAKAEKFPDLATT 115

Db 769 QIALAASKSAASSSSAASAAIV--ASSNAASEASSAAAAVSNADANSAAAYDSYASE 826

Qy 116 ASSA-----FENTPFAASVLQYMQPAINKGDWLATPLKPLTPLISGALSAMDQVG 167
Db 827 ASAASANDSSGVATASFAASSAAAAAM-----SAALSTAQVAAK 865
Qy 168 TKMDRARGDLHLVLTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFESALNVVR--T 225
Db 866 VAYSDAAAAG-----SAAVAVASAAQSDSKKKQATATARSQAALDDLKIKSLT 913
Qy 226 VLAPALASRPSVQ-----AVDFGVSTAGGIVANAGFGDRMLSVQSRDQLRGGA 274
Db 914 DYASGASSASAEAGQASTATSAVASAASSSASEAGSYAHQAGSSADAVGQSGSAAQHAS 973
Qy 275 FVLGMKKEKPKAALSEETDMLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRS 334
Db 974 TAASAASSYPKDSGIQSLASQAASEAAKASSNAATSA-----AAGFSAASDASEQAKT 1029
Qy 335 LVSATSLTKNGALAGGYAGVSKLQKWATKNITDSATKAAVSQLSNLVGVSVPAGWTTA 394
Db 1030 AASADVAVASSAASATANSASAAA-----SATKAGDSKAA-----AGFSSA 1069
Qy 395 GLATDPAVKKAESPIQKVKTAS--STTSYVADQTVKLAKTVKDMSGEAISSGTGSLRS 452
Db 1070 ASAAASAKGAEEAVASEAASAAASDDSVASSAASAAAGFDKAAASAAEGAAASAAAS 1129
Qy 453 TVNNLRHRSAPEDIIEGGIS 473
Db 1130 AAQGTGGASSSASEAGQAS 1150

RESULT 5

AAE00701

ID AAE00701 standard; Protein; 2123 AA.

AC AAE00701;

DT 02-JUL-2001 (first entry)

DE Moraxella catarrhalis outer membrane protein-106 (OMP106).

XX Outer membrane protein-106; OMP106; haemagglutination; vaccine; bacterial infection; immunogen; cytotoxic; antibiotic; passive immunisation.

XX Moraxella catarrhalis.

XX US6214981-B1.

XX 10-APR-2001.

XX 12-NOV-1997; 97US-0968685.

XX 03-MAY-1996; 96US-0642712.

XX (ANTE-) ANTEX BIOLOGICS INC.

XX Tucker K, Plosila L, Tillman UP;

XX WPI; 2001-281002/29.

XX N-PSDB; AAD04029.

XX Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane protein-106 polypeptide, useful for diagnosis of bacterial infections and as vaccine against Moraxella catarrhalis infection of mammals -

XX Claim 7; Column 53-64; 49pp; English.

XX The present sequence is haemagglutinating Moraxella catarrhalis outer membrane protein-106 (OMP106). The OMP106 is used as a therapeutic and prophylactic vaccine against M. catarrhalis infections of mammals. It is used for diagnosis of bacterial infections and as reagents for clinical or medical diagnosis of M. catarrhalis infections and for scientific research on the properties of pathogenicity, virulence and


```

Db 719 TAGLEAGTRAAVASRVLKLTALRGLLVSTLVGGFAALGWALESLSFAEAKKAKODFE 778
Qy 406 -----BSFIO-----DKVKSTASSTSYVADQ---TVKLAKT-----VK 436
Db 779 QSQQTNEAITTKNDSTDKLIQYKELQKVKEBSRLTSDEQEYLVQVTOQLAQTPPALVK 838
Qy 437 --DMSGEAISSTGASLRSTVNNLR 458
Db 839 GYDSQGNAILTKNKELEAIENK 862

RESULT 7
ABB71736
ID ABB71736 standard; Protein; 1229 AA.
XX
AC ABB71736;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 42000.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL15839.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 42000; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABE57737-ABE72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1229 AA;

Query Match 5.7%; Score 138; DB 22; Length 1229;
Best Local Similarity 20.9%; Pred. NO. 0.022;
Matches 107; Conservative 73; Mismatches 205; Indels 128; Gaps 21;

Qy 11 PCVAMESFRTASDASLASS--VRSVSTSCNDLQAITDYLKHVFAAHRFSVIGSPDER 68
Db 166 PGITVATYTPGYSYSSAPGQISKVATYSSPSLSVPYAPPVTSKVETYS---SP--- 219
Qy 69 DAALAHNQID--ALVETRNRLYSEGETPATI-----AETFAKAEKFD--- 110

```

```

Db 220 --AVYTKTTPGYSKVETYSSPGYSYGOISPGISRIATYSPSYVSAPTIAKVSTYSAPS 277
Qy 111 RLATTAS-SAFENTPFAA---SVLOYMOPAINKGWMLATPLKPLTPLIS-----GA 158
Db 278 VKLATTSSLLSHGTGYSAQYAPSIITKYQSVDSVSHQYISKPIVAAAPAITKVAASYGGT 337
Qy 159 LSGAMDQ-----VGTKWMDRARGDLHLSTSP--DKLHDAMAVSVKHSHPALGRQVVD 210
Db 338 ASGALSHQYVSQPAIAKVSTYVAPTATYSSGPAISKLSYSTYGAS---GSGAVSHQYVSK 394
Qy 211 GIAVQTFESALNVVTRTVLAPALAS---RPSVQGVDFGVSTAGGLVANAGFGDRMLSVQSR 267
Db 395 PAVAIAAPAVAKVATYAAPAISSYSTGPAISKVASYAAPTSTYSSGYGVS----- 446
Qy 268 DQLRGGAFLVMKDKPKAALSEETDMLDAYKAI-KSASYSGAALNAGKRMAGLPL----- 322
Db 447 ---SGGAVSHQYVSKPAVAIS-----AAPAIKAVATYAAPAIST---YAAAPVVTKV 493
Qy 323 -----DVATDGLKAVRSLSVTSLSLTNGNLALAGGAGV 355
Db 494 ATGYGSGSGYSGGAVSHQYVSKPAVAKVATYAAPAISTYSAAPAVTKIATSYGSGHGA 553
Qy 356 SKLQKMATKNIITDSATKAAVSQLSNLVGSGVFAGMTTAGLATDPAVKAESFTQDKVKS 415
Db 554 VSHQYVSKPAVAISAAPA-----IAKVATYASPAISTYATAPVVKVATYAAPSIAT 605
Qy 416 TASSTSYVADQTVKLAKT-----VKDMSGEAIS 444
Db 606 YSSAPA-----LAKVSYQAADVSHQYVIS 629

RESULT 8
AAB76536
ID AAB76536 standard; Protein; 1277 AA.
XX
AC AAB76536;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein SEQ ID NO:54.
XX
KW Corynebacterium glutamicum; brevbacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100805-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00926.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.

```


PR	03-SEP-1999;	99DE-1042078.
PR	03-SEP-1999;	99DE-1042079.
PR	03-SEP-1999;	99DE-1042088.
XX	(BADI) BASF AG.	
PA	Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;	
PI	WPI; 2001-071486/08.	
DR	N-PSDB; AAF67767.	
XX	Corynebacterium glutamicum nucleic acids encoding membrane construction	
PT	and membrane transport proteins or their portions, useful for typing or	
PT	identifying C. glutamicum or related bacteria, and as markers for	
PT	transformation -	
XX	Claim 20; Page 224-231; 1119pp; English.	
PS		
XX	AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane	
CC	construction and membrane transport (MCT) proteins given in AAB76510 to	
CC	AAB76847. The MCT nucleic acids and proteins are useful in the	
CC	identification of microorganisms which can be used to produce fine	
CC	chemicals, for modulating fine chemical production in C. glutamicum or	
CC	related bacteria (e.g. Brevibacterium lactofermentum), the typing or	
CC	identification of C. glutamicum or related bacteria, as reference points	
CC	for mapping C. glutamicum genome, and as markers for transformation.	
CC	AAF68082 and AAF68082 represent sequencing primers which are used in an	
CC	example from the present invention.	
XX		
SQ	Sequence 2012 AA;	
	Query Match 5.7%; Score 138; DB 22; Length 1012;	
	Best Local Similarity 21.7%; Pred. No. 0.044;	
	Matches 119; Conservative 85; Mismatches 219; Indels 126; Gaps 23;	
QY	7 AQOPPGVAMESFTASDASLASSSVSSTTSCRDLOAITDKLKHVFVAHRFSVIGSPD 66	
Db	995 AHAPDLVLGRAPFAVFAAKSAVIPGTDSASVEGMLSLVLEHHI-----VLKSDV 1047	
QY	67 ERDAALAHNEQIDALVETRANRLY-----SEGETPATIAETFA-KAEKFDRLATIAS 117	
Db	1048 PTDGALKVKSATADEVDITGLRILIVRAIEADAGNLIATLAERFAIRGKNNAVARTNT 1107	
QY	118 SAFE---NTPFAAASVLQYNQP-----AINKGDMLATPLKPLTPL-----ISGALSAMD 164	
Db	1108 SALPVTDTSRARAVATTVAPESMEPFVISED-----RNPHVSDVAASLAGLP 1159	
QY	165 QVGTKMDRARGDLHYLSPPKLHDAMAVSVKRH-----SPALGRQVVDMGI-----AVOT 216	
Db	1160 VIVHGMMWTSAGIELIAGAANDFEIQTPAAKVVEYTATMLAPVLGEIEFFSVERSAVDN 1219	
QY	217 FSALNVRTVLAPALARSPOCAQDFGVSTAGGLVA--NAGFGDRMLSVOQRDLRGGA 274	
Db	1220 RPKGGEVRTVTA-----TVNGNL---VLTATVVAAPTFTYFPGGIOIQSG----- 1263	
QY	275 FVLGMKDKEPKAALSBETOWLDAYKAIKASYSGAALNAGKRMAGLPIDVATDGLKAVR- 333	
Db	1264 --MGMEARRNSQARAIWDADAHTRNK-----LGFSIVEIVENNPREVTVAGEKFPHP 1315	
QY	334 -SLVATSFLTKNGLALAGGYAGVSKLQMKATKNITDSATKAASVQLSNLVGSVGFAGWT 392	
Db	1316 DGVLYLTQTFQVMATL-GVAQIAEMREAHALNORAYFAGHSVGEYNALAAVAGVLSLES 1374	
QY	393 T-----AGLATDPKAKBASFTQ-----DKVKSTASTTTSVADOITVKLAKTYKD 437	
Db	1375 VLIEIVRRGLTHRLVRDENGSLSNVALAALRNKGLTADNVFDYA-----SYSE 1426	
QY	438 MSCE-----ATSSGASLRSTVNHLRHRSAPAEDIEE--CGISAF---SRS 478	
Db	1427 ASGEFLEIVNNYLAGLOYAVAGTOAGL-----AALRADVENRAPGQAFILIPGI 1476	
QY	479 ETPFQLRL 487	

Db	1477 DVFPFHSSKL 1485
RESULT 10	
AAG92485	
ID	AAG92485 standard; Protein; 2993 AA.
XX	AAG92485;
AC	AAG92485;
XX	26-SEP-2001 (first entry)
DT	
XX	C glutamicum protein fragment SEQ ID NO: 6239.
DE	Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX	organic acid synthesis.
KW	Corynebacterium glutamicum.
XX	EP1108790-A2.
PN	20-JUN-2001.
XX	18-DEC-2000; 2000EP-0127688.
PF	16-DEC-1999; 99JP-0377484.
XX	07-APR-2000; 2000JP-0159162.
PR	03-AUG-2000; 2000JP-0280988.
PR	(KYOW) KYOWA HAKKO KOGYO KK.
XX	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
PI	WPI: 2001-376931/40.
XX	N-PSDB; AAG7704.
DR	Novel polynucleotides derived from Coryneform bacteria, for identifying
XX	mutation point of a gene, measuring expression of a gene, analysing
PT	expression profile or pattern of a gene and identifying homologous gene
PT	-
PT	Claim 17; SEQ ID NO: 6239; 246pp + Sequence Listing; English.
XX	The present invention provides a number of nucleotide and protein
CC	sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC	are useful for identifying the mutation point of a gene derived from a
CC	mutant of coryneform bacterium, measuring expression amount and
CC	analysing the expression profile or expression pattern of a gene derived
CC	from Coryneform bacterium, and identifying a homologue of a gene derived
CC	from coryneform bacterium. Coryneform bacteria are useful for producing
CC	amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC	particularly L-lysine. The present sequence is a protein described
CC	in the exemplification of the invention.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from the
CC	European Patent Office.
XX	Sequence 2993 AA;
SQ	
	Query Match 5.7%; Score 138; DB 22; Length 2993;
	Best Local Similarity 21.7%; Pred. NO. 0.076;
	Matches 119; Conservative 85; Mismatches 219; Indels 126; Gaps 23;
QY	7 AQOPPGVAMESFTASDASLASSSVSSTTSCRDLOAITDKLKHVFVAHRFSVIGSPD 66
Db	1025 AHAPDLVLGRAPFAVFAAKSAVIPGTDSASVEGMLSLVLEHHI-----VLKSDV 1077
QY	67 ERDAALAHNEQIDALVETRANRLY-----SEGETPATIAETFA-KAEKFDRLATIAS 117
Db	1078 PTDGALKVKSATADEVDITGLRILIVRAIEADAGNLIATLAERFAIRGKNNAVARTNT 1137
QY	118 SAFE---NTPFAAASVLQYNQP-----AINKGDMLATPLKPLTPL-----ISGALSAMD 164

Db 1138 SALTPTVDTSRSAVAATVVAPESMRPFVATISGD-----RNPVHSDVAASLACPG 1189
 QY 165 QVETKMDRARGDLHYLSTSPDKLHDMAVSVKRN-----SPALGRQVDMGI---AVQT 216
 Db 1190 VIVHGMWTSALIGELIAGAANFDEQITPAAKVVEYATMTLAPVLPGSEIEFSVERSAVDN 1249
 QY 217 FSLANVTRTLAPLAPLAPSPVQGVNDGVSTAGGLVA--NAGFDBRMLSVOSRQDLRGGA 274
 Db 1250 RPGMGVRRITVTA-----TVNGNL-----VLTATVVAAPSTFVAFPCQGIQSOG----- 1293
 QY 275 FVLGMDKEPKPALSEETDMLDAVKAIKSASYSGALNAGKRMGLPLDVAITDGLKAVR- 333
 Db 1294 --MGMERARRSQARAAWDRADATRNK-----LGFSEIYEIVNNPREVTVAKEKFFHP 1345
 QY 334 -SLVSAISLTKNGLAGVAVSKLOKMAKNTDTSATKAASQSLNLSVGVFPAWMT 392
 Db 1346 DGVLYLQFQVGMATL-GVAQIEMREAHMLNQRAVFAHGSVEGVNALAAVAVLSLES 1404
 QY 393 T-----AGLATDPAVKKASFIQ-----DKVSTASSTTSYVADQTVKLAKTVKD 437
 Db 1405 VLEIVRRGLTMHRLVDRDNGLSNVALAALRPKMGILTADNVDYVA-----SVSE 1456
 QY 438 MSGE-----AISTGASLSTVTNNLRHSAPRADIEE--CGISAF---SRS 478
 Db 1457 ASGEFLEIVNNMLAGLOYAVAGTQAGL-----AALRADVENRAPQGRAFILIPGI 1506
 QY 479 ETPFQRL 487
 Db 1507 DVPFHSEKL 1515

RESULT 11
 AAR48993
 ID AAR48993 standard; Protein; 1026 AA.

XX AAR48993;

XX 14-SEP-1994 (first entry)

XX resA S-lyase protein.

KM C. crescentus; resA; paracrystalline; S-layer; protein; heterologous;
 KM cellulase; xylase; metallochionein; restriction site;
 KM reading frame; fusion protein; bioreactor; toxic metal; sewage;
 KM waste water; wood pulp suspension; cell surface; vaccine; fish.

XX Caulobacter crescentus.

XX CA2090549-A.

XX 10-DEC-1993.

XX 26-FEB-1993; 93CA-2090549.

XX 09-JUN-1992; 92US-089367.

XX (UYBR-) UNIV BRITISH COLUMBIA.

XX Bingle WH, Smit J;

XX WPI; 1994-066249/09.

XX N-PSDB; AAO57972.

XX Prod. of heterologous polypeptides in bacteria, partic.

XX Caulobacter - by expression of a fusion prod. of the polypeptide

XX sequence and a bacterial S-layer protein gene

XX Claim 17; Fig 6; 27pp; English.

CC This sequence is encoded by the C. crescentus resA gene and represents
 CC the paracrystalline S-layer protein. The resA gene was used in the
 CC production of the heterologous protein of the invention. The
 CC heterologous protein is produced by cloning a polypeptide coding

CC sequence, eg. cellulase, xylase or a metallochionein, into a
 CC restriction site within the resA gene which preserves the resA reading
 CC frame and expressing the fusion sequence in Caulobacter. This S-
 CC layer protein bacterial system can be used in bioreactors, eg. to
 CC bind toxic metals in sewage waste water etc. or for the treatment of
 CC wood pulp suspensions. The system can be used to produce heterologous
 CC proteins at the cell surface for use in vaccines, partic. fish
 CC vaccines. The S-layer protein is synthesised in large quantities and
 CC has a general repetitive sequence, permitting the synthesis of large
 CC amounts of heterologous protein as a fusion product and presentation at
 CC the cell surface.

XX Sequence 1026 AA;

XX Query Match 5.7%; Score 136; DB 15; Length 1026;

XX Best local Similarity 21.2%; Pred. No. 0.025;

XX Matches 102; Conservative 75; Mismatches 165; Indels 140; Gaps 23;

QY 22 SDAASLASSVSVSTTSCRDQAITDYLKHHVFAHRFSVIGSPDBDALAHNEQIDAL 81

Db 42 SDAALTNLTKLVNSTAVAIQT-----YQFTGVAPSAAG-----LDPL 81

QY 82 VETRRAN-----RLYSEGERPATIAETPAKAEK-----DRLATTASAPENTPPAA-- 128

Db 82 VDETNTNDLNDAYYSK-----FAQENRPFINSINLATGAGG--ATAFAAAAYTG 129

QY 129 -SVLYQVQPAINK--GDMLATPLKPLTPLISGALSGAMPDQVTKMDRARGDLHYLSTSP 185

Db 130 VSYAQVYATAYDKIIGNAVT-----AAGVDVAANAVALPSQANIDYLT-- 173

QY 186 DKLHDMANVSVKHSPALGRQVNDMGIAVQTFSAALNVRTVL--APALASRPSVQAVD 242

Db 174 -----AFVRANTPFTAAIDIDLAVKALIGTIILNAAVSGIGYATATAMINDLSD 225

QY 243 FGVST--AGGL-----VANAQFDRMLSVSR--DQRG-----GAFVLGMDK 261

Db 226 GALSTDNAAAGVNLFTAYPPSSGVSSTLSLTRTDTLTGTANNDFVAAGEVAGAAITLVGD 285

QY 282 KEPKALSEETDMLDAVKAIKSASYSGAALNAGKRMAGPLDVAITDGLKAVRSILVSATSL 341

Db 286 TLGGAGCTDVLNMQA-----AAVTA-----LPTGVTIGIEIM--NTSGAAI 327

QY 342 TKNGLALAGYAVSKL-----QKXATKNTDTSATKAASQSLNLSVGS 384

Db 328 TLN---TSSGVTLTLNTNTSGAQTVAAGQNTLTATTAQAANNVAVDGGANVTVAS 384

QY 385 VGVFAGTTTGLATDPAVKKASFIQDKYKSTASSTTSYVADQTVKLAKTVKMSGEAIS 444

Db 385 TGVTSGITTVG-----ANSNAGSTVSVANSSTTTTGAIA-VTGGTAVTVQTAGNANV 438

QY 445 ST 446

Db 439 TT 440

XX RESULT 12

XX AAM37490

XX ID AAM37490 standard; Protein; 1026 AA.

XX AAM37490;

XX 20-APR-1998 (first entry)

XX Caulobacter crescentus S-layer resA protein.

XX S-layer; resA gene; Caulobacter; vaccine; antigenic; ligand; enzyme;

XX metallochionein; heavy metal; water; sewage; xylanase; cellulase;

XX wood pulping.

OS Caulobacter crescentus.
 XX CA2090549-A.
 XX WPI; 1994-066249-09.
 XX N-PSDB; AAO57972.
 XX Prod. of heterologous polypeptides in bacteria, partic.
 XX Caulobacter - by expression of a fusion prod. of the polypeptide
 XX sequence and a bacterial S-layer protein gene
 XX Claim 17; Fig 6; 27pp; English.

PD 18-SEP-1997.
XX 10-MAR-1997; 97WO-CA00167.
XX 12-MAR-1996; 96US-0614377.
XX (UYBR-) UNIV BRITISH COLUMBIA.
XX Bingle WH, Nomellini JF, Smit J;
XX WPI: 1997-470880/43.
XX N-PSDB; AAV01866.
XX New DNA containing sequence for C-terminal region of Caulobacter
PT S-layer protein - expressed as fusion proteins containing antigenic
PT peptides in Caulobacter, useful as live vaccines
XX Example 2; Fig 6; 58pp; English.
XX The present sequence represents a Caulobacter S-layer protein used in an
CC example of the present invention. A new DNA construct has been developed
CC which contains at least one restriction site for insertion of DNA
CC upstream of DNA encoding a C-terminal region of at least the last 82
CC amino acids (aa) of Caulobacter S-layer protein. Caulobacter containing
CC the DNA constructs above additionally containing a sequence encoding a
CC heterologous polypeptide, are particularly useful in live vaccines
CC (where the heterologous polypeptide is an antigen). They can also be
CC used for production of e.g. ligands, enzymes or other proteins, e.g.
CC metallotheins to remove heavy metals from water or sewage, or
CC xylanase or cellulase for use in wood pulping. All known Caulobacter
CC strains are harmless, and stable in outdoor environments, including
CC water (so suitable for vaccinating fish) or soil. They are well suited
CC for growing in biofilm reactors and produce S-layer proteins, which is
CC an ideal system for presentation of antigens, at high level.
XX
XX Sequence 1026 AA;
XX
XX Query Match 5.5%; Score 132; DB 18; Length 1026;
XX Best Local Similarity 21.0%; Pred. No. 0.054;
XX Matches 101; Conservative 75; Mismatches 166; Indels 140; Gaps 23;
XX
Qy 22 SDASLASSSVRSVTSCTDLQAITDYKHHVFAHRSFVSGSPDERDAALAHNEQIDAL 81
Db 42 SDAALATNTLVNSTTAVAIQT-----YQFTGVAFSAAG-----LDLFL 81
Qy 82 VETRAN-----RLYSEGETPATIAETFAKEKF-----DRLATTASSAFENTPFAAA--- 128
Db 82 VDSITNTDNDAYISK-----FAQENRFNFESINLATGAGG--ATAFAAATG 129
Qy 129 -SVLYQMOPAINK--GDMLATPLKPLTPLISGALSGAMDQVGTNMDRAGDLHYLTSP 185
Db 130 VSYAQTAVATYDKIIGNAVAT-----AAGVDVAAVAFLSRQANIDYLT--- 173
Qy 186 DKLHDAMAVSKRHPALGRQVDMGIAVQFSAALNVRTVL---APALASRPSVQGAVD 242
Db 174 -----AFVRANTPFTAAADIDILAVKAALIGTILNATVSGIGYATATAAMINDLSD 225
Qy 243 FGVSF--AGGL-----VANAGFGDRMLSVOS--RDQLRG-----GAFVLGMDK 281
Db 226 GALSITDNRAGVNLFTAYPSGSGSTLSLTITGTDLTGTANNDFVAGEVAGATLTGVD 285
Qy 282 REPKAALSEETDLDAYKAIKSASYSAGALNAGKRMAGLPDLDVATDGLKAVRSLVSATSL 341
Db 286 TLGGGAGTDLVNWQA-----AAVTA-----LPTGVTTISGIETM-NVTSGAAI 327
Qy 342 TKNGLALAGGYAGVSKL-----QKWKNTITDSATKAQVLSQLNL-VGS 384
Db 328 TLN-----TSSGVTLTALNTNTSGAAQTVTAGAGQNLTAATAQAANNVAVDGRANVTAS 384
Qy 385 VGVFAGWTTAGLATDPVAKKAEFSIQDKVKSTASTSTTSYADQTVKLTAKTKDMSGEAIS 444
Db 385 TGVTSGETTVG-----ANSAASGTVSVSVANSSTTTTGAIA-VTGGTAVTVAQTAGNAV 438

Qy 445 ST 446
Db 439 TT 440
RESULT 13
AAAY44757
ID AAY44757 standard; Protein; 1026 AA.
XX AC AAY44757;
XX 04-MAY-2000 (first entry)
XX Caulobacter crescentus surface layer protein.
XX Surface layer protein; S-layer secretion signal; antibiotic; vaccine;
XX recombinant fusion protein cleavage; enzyme; protein polymer;
XX antibacterial enzyme; foodstuff.
XX Caulobacter crescentus.
XX Key Location/Qualifiers
XX Cleavage-site 692..693
XX /note= "Asp-Pro dipeptide present in S-layer secretion
XX signal sequence. It is a site where a fusion
XX protein comprising a target protein and the secretion
XX signal is cleaved"
XX WO200004170-A1.
XX 27-JAN-2000.
XX 14-JUL-1999; 99WO-CA00637.
XX 14-JUL-1998; 98CA-2237704.
XX (UYBR-) UNIV BRITISH COLUMBIA.
XX Smit J;
XX WPI: 2000-182434/16.
XX N-PSDB; AAZ50079.
XX Cleavage of Caulobacter produced recombinant fusion proteins useful for
XX producing vaccine peptides -
XX Example 1; Pages 21-23; 33pp; English.
XX The patent discloses a method for cleaving a recombinant fusion protein
XX which is produced by Caulobacter and consists of Caulobacter surface
XX layer (S-layer) protein (containing the C-terminal secretion signal) and
XX a target protein heterologous to Caulobacter. The cleavage of target
XX protein from the S-layer protein is carried out under mild acid
XX conditions so that cleavage occurs at aspartate-
XX proline dipeptide site without solubilising the protein.
XX The cleavage is accomplished while the fusion protein is in
XX an insoluble aggregate form which facilitates purification of the
XX protein. The method is useful for producing pure proteins including
XX recombinant human and animal therapeutic antibiotic and vaccine peptides,
XX enzymes, protein polymers, and antibacterial enzymes for foodstuffs.
XX The present sequence is a S-layer protein from C. crescentus.
XX The S-layer secretion signal, corresponding to the C-terminal portion of
XX the protein from amino acid 690 onwards, is fused with a target sequence
XX for construction of a recombinant fusion construct which is expressed in
XX Caulobacter.
XX
XX Sequence 1026 AA;
XX
XX Query Match 5.5%; Score 132; DB 21; Length 1026;
XX Best Local Similarity 21.0%; Pred. No. 0.054;
XX Matches 101; Conservative 75; Mismatches 166; Indels 140; Gaps 23;
XX
Qy 22 SDASLASSSVRSVTSCTDLQAITDYKHHVFAHRSFVSGSPDERDAALAHNEQIDAL 81

```

Db 42 SDAALNTLTKLVNSTVAVAIQT-----YQFFGVAPSAAG-----LDFL 81
QY 82 VETRAAN-----RLYSEGETPATIAETPAKAEKF-----DLATASSAFENTPPAA--- 128
Db 82 VDSSTNTNDLNDAYYSK-----FAQENRFINFSINLATGAGG--ATAFAAAVTG 129
QY 129 -SVLYQWQAPINK--GDMLATPLKPLPLPLSGALSGAMDQVGTQKMDRARGDLHYLSTSP 185
Db 130 VSVAYQVATATVDKLIQNAVAIT-----AAGDVAAVAAFLSRQANDIYLT--- 173
QY 186 DKLHDMAAVSVKRSPLAGROVDMGIAVOTFSALNVRTYL---APALASRPVQGAVD 242
Db 174 -----AFVRANTPPTAAADIDLAVKALITILMAATVSGIGYATATPAAIMINDLS 225
QY 243 FGVSF--AGGL-----VANAGFDRMLSVQS--RDQLG-----GAVLGMD 281
Db 226 GALSTDNAAGVNLFTAVPSSGVSSTLSTTGTTLGTANNDTFVAGEVAGAAATLTVD 285
QY 282 KEPKALSEETDMLDAVKAIKSASVSGAALNAGKRMAGLPUDVATDGLKAVRSIVSATSL 341
Db 286 TLSSGACTVDLNMVQA-----AAVTA-----LPTGVITISGIETM--NTVSGAAI 327
QY 342 TKXGLALAGYAGYSK-----OKMATKNITDSATRAAVSOLSNL--VGS 384
Db 328 TLN---TSSGVTGLTALNTVTSGAAGQVTAGAGQNLTAATTAQAANNVAVDGRANVTVAS 384
QY 385 VGVFAGTGTAGLATDPAVKKAEFIDDKVSTASTSTSYVADQVVKLAKTVKMGSGEALS 444
Db 385 TGVTSSTTTVG-----ANSASGTVSVSVANSSTTTGAIL-VTGSTAVTVAAQTAGNAVN 438
QY 445 ST 446
Db 439 TT 440

RESULT 14
AAU34139
ID AAU34139 standard; Protein; 2368 AA.
XX
AC AAU34139;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #415.
XX
KM Antisense; prokaryotic cellular proliferation protein;
XX
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
XX
PR 23-MAY-2000; 2000US-206848P.
XX
PR 26-MAY-2000; 2000US-207727P.
XX
PR 23-OCT-2000; 2000US-242578P.
XX
PR 27-NOV-2000; 2000US-253625P.
XX
PR 22-DEC-2000; 2000US-257931P.
XX
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zykkind JW, Wall D, Trawick JD, Carr GJ;
XX
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX
DR N-PsDB; AAS51998.
XX

```

```

PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
PS Example 3; Seq ID No 5635; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2368 AA;
XX
Query Match 5.4%; Score 130.5; DB 22; Length 2368;
Best Local Similarity 20.2%; Fred. No. 0.23;
Matches 99; Conservative 64; Mismatches 223; Indels 105; Gaps 17;
QY 4 NOSAQCPGVAVMESFRTAS-----DASLASSVSVSTTSCRDQATRDYLKHVFA 56
Db 702 NASLQDEKVANDKIGIEKIEKAIKDIDATTAQVEAIKTAINDINQTPATTAQAAL 761
QY 57 HRSFVIGSPDERDAI---AHNEQIDALVETRAMRLYSEGETPATIAETPAKAEKEDRLA 113
Db 762 EEFDEVVQAQIDQAPLNPTTNEVAEIE---RINAQVSGVKAIEATTAQDLERVK 817
QY 114 TTASSAFENTPPAAASVLQWQAPINKGDMLATPLKPLPLPLSGALSGAMDQVGTQKMDR 173
Db 818 NEBISKIEN---ITDSTQTKMDAYNEVKQAATARKTONATVSNATN---EEV----- 863
QY 174 ARGDLHYLSTPDKLHDMAAVSVKRSPLAGROVDMGIAVOTFSALNVRTYLAPALAS 233
Db 864 AEAADAVERAQAQGLHDIOVVKSKQEVADTKSKVLDKINAIQT----- 906
QY 234 RPSVQGAVDPEGVSTAGLVANNAFGDRMLSVOGRDQLRGAFVLGKMKDKPKALSEETD 293
Db 907 QAKVKPAAADTEVENA-----YNTRKQEIQNSN-----ASTTEKQAAYTE--- 946
QY 294 WLDAYKAIKSASVSGAALNAGKRMAGLPUDVAT---DGLKAVNSLVYSLSLTNGGLALAG 350
Db 947 -LDTKQGEARTNLDAANTMS-----DVTAKONGIAIAIQVQAATTKKSDARAIEA 996
QY 351 GYAGVSKL-----OKMATKNITDSATRAAVSOLSNLVGSVGFAGWT---A 394
Db 997 QKASEKRTAIEAMNDSTTEEQAAKQVDQAVVTTAAADIDNAAANTVDVNAKTINERTIA 1056
QY 395 GLATDPAVK-KAESFIQDKVKSF-----ASSTTSYVADQTVKLIATVKMGSGEALS 444
Db 1057 AITPPDANVKEPTAKQALADKVAQOETAIDANNNGATTEBKAAKQGVQTEKTTADTAIDG-A 1115
QY 445 STGASLRSTVN 455
Db 1116 HTNAVEVAAKN 1126

RESULT 15
AAU36796
ID AAU36796 standard; Protein; 2368 AA.
XX
AC AAU36796;
XX

```

XX 14-FEB-2002 (first entry)
 XX Staphylococcus aureus cellular proliferation protein #966.
 XX DE
 XX Staphylococcus aureus cellular proliferation protein;
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX OS
 XX Staphylococcus aureus.
 XX PN W0200170955-A2.
 XX PD 27-SEP-2001.
 XX PF 21-MAR-2001; 2001WO-US09180.
 XX PR 21-MAR-2000; 2000US-191078P.
 XX PR 23-MAY-2000; 2000US-206848P.
 XX PR 26-MAY-2000; 2000US-207727P.
 XX PR 23-OCT-2000; 2000US-242578P.
 XX PR 27-NOV-2000; 2000US-253625P.
 XX PR 22-DEC-2000; 2000US-257931P.
 XX PR 16-FEB-2001; 2001US-269308P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX WPI; 2001-611495/70.
 DR N-PSDB; AAS54655.
 XX New polynucleotides for the identification and development of
 XX antibiotics, comprise sequences of antisense nucleic acids -
 XX Example 3; Seq ID No 12389; 511pp; English.
 XX The invention relates to antisense inhibitors of genes essential to
 XX prokaryotic cellular proliferation, their use in identifying the
 XX genes, their use in the discovery of novel antibiotics, the essential
 XX genes themselves and the encoded proteins. The prokaryotes used are
 XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 XX invention is also useful for the identification of potential new targets
 XX for antibiotic development. The antisense nucleic acids can also be used
 XX to identify proteins used in proliferation, to express these proteins,
 XX and to obtain antibodies capable of binding to the expressed proteins.
 XX The proteins can be used to screen compounds in rational drug discovery
 XX programmes. The antisense nucleic acid sequence is also useful to screen
 XX for homologous nucleic acids which are required for cell proliferation in
 XX a wide variety of organisms. The present sequence represents an
 XX essential prokaryotic cellular proliferation protein.
 XX Note: The sequence data for this patent did not form part
 XX of the printed specification, but was obtained in electronic
 XX format directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 2368 AA;

Query Match 5.4%; Score 130.5; DB 22; Length 2368;
 Best Local Similarity 20.2%; Fred. No. 0.23;
 Matches 99; Conservative 64; Mismatches 223; Indels 105; Gaps 17;

QY 4 NOSAQQPPGVAMESPTAS-----DASLASSSVRSVSTTSCRDLOAITDYLKHHVFAA 56
 DB 702 NASLQDEKDVANDKIGKIEKTKAIDDAATNAQVEAIKTKAINDINQTPATTAKAAL 761
 QY 57 HRFVSVIGSPDERDAAL---AHNEQIDALVETRANRKYSEGETPATIAETFAKAEKFDRLA 113
 DB 762 EEFDEVVQAQIDQAPLNPDTTNEEVAEAE----RINAAKVSGVKAIEATTTAQDLERVK 817
 QY 114 TTASSAFENTPFAASVQYMQPAINKGDWLATPLKPLTPLISGALSGAMDOQVGTMMDR 173

Db 818 NEBISKIEN-----ITDSTQTKMDAYNEVKAQATARTKQNTATVSNATN---EEV----- 863
 QY 174 ARGDLHYLSTSPDKLHDAMAVSVKRRHSPALGRQVDMGIADVQTFSSALNVVTVLAPALAS 233
 Db 864 AEADAAVEAAKQGLHDIQVVKSKQEVADTKSVLDKINAIQT----- 906
 QY 234 RPSVQGNVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFLVGMKDKPKAALSEETD 293
 Db 907 QAKVPAADTEVENA-----YNTKQBIQNSN-----ASTTEKQAAYTE--- 946
 QY 294 WLDAYKAIKSASYSGAALNAGKRMAGLPLDVAT---DGLKAVRSLVSATSLTKNGLALAG 350
 Db 947 -LDYKQEQARTNLDAAANTS-----DVTAKNGIAIINQVQAATTKKSDAKAEIA 996
 QY 351 GYAGVSKL-----OKMATKNIITDSATKAAVSQSLNVLVSGVGFAGWTT---A 394
 Db 997 QKASERKTAJEAAMNDSTTEEQAAKQKVDQAVVTANADIDNAAAANTDVDNAKTNEATIA 1056
 QY 395 GLATDPAVK-KAESFIQDKVKST-----ASSTTSYVADQTVKLAKTVKMSGEAIS 444
 Db 1057 AITPDANVVKPTAKQAIADKVQAQETAIDANNNGATTTEKAAAKQOVQTEKTTADTAIDG-A 1115
 QY 445 STGASLRSTVN 455
 Db 1116 HTNAEVEAAKN 1126

Search completed: January 2, 2003, 15:19:00
 Job time : 42.5355 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:16:50 ; Search time 30.0308 Seconds
(without alignments)
3341.397 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407
Sequence: 1 MHINQSAQPPGVAMSEFRT.....EEGGSIAFSRSETPQLRRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1869.5	77.7	486	Q9JP34	Q9JP34 pseudomonas
2	1591.5	66.1	484	Q8RP03	Q8RP03 pseudomonas
3	460	19.1	518	Q8XQF0	Q8XQF0 ralsstonia s
4	149.5	6.2	2155	Q8UPP9	Q8UPP9 agrobacteri
5	149	6.2	1009	Q8UWU1	Q8UWU1 agrobacteri
6	146	6.1	2089	Q92K98	Q92K98 rhizobium m
7	143.5	6.0	2055	Q85472	Q85472 abiotrophia
8	141	5.9	536	Q9HFA2	Q9HFA2 halobacteri
9	140.5	5.8	642	Q8XSW6	Q8XSW6 ralsstonia s
10	140.5	5.8	1983	Q93JN9	Q93JN9 staphylococ
11	140.5	5.8	2186	Q997B0	Q997B0 staphylococ
12	140	5.8	503	Q9FC05	Q9FC05 pseudomonas
13	140	5.8	2285	Q64046	Q64046 bacterioph
14	140	5.8	2285	Q31976	Q31976 bacillus su
15	139.5	5.8	491	Q93GTL	Q93GTL campylobact
16	139.5	5.8	1795	Q9LCO9	Q9LCO9 staphylococ

17	138.5	5.8	491	2	Q8RTY4	Q8RTY4 campylobact
18	138.5	5.8	548	2	Q9RA74	Q9RA74 streptococ
19	138	5.7	1229	5	Q9VW05	Q9VW05 dirosophila
20	138	5.7	1731	16	Q8UWU4	Q8UWU4 agrobacteri
21	137.5	5.7	2283	2	Q8VQ99	Q8VQ99 staphylococ
22	137	5.7	1545	16	Q9PD01	Q9PD01 streptomyc
23	135	5.6	973	16	Q8XDQ4	Q8XDQ4 escherichia
24	135	5.6	6077	12	Q8VAP1	Q8VAP1 white spot
25	135	5.6	6077	12	Q8QTB7	Q8QTB7 white spot
26	134.5	5.6	1713	3	Q8TGE1	Q8TGE1 saccharomyc
27	134.5	5.6	2016	5	Q9BIT0	Q9BIT0 plectenrys
28	134.5	5.6	2178	2	Q9KMR3	Q9KMR3 streptococ
29	134	5.6	2535	16	Q8ZDR6	Q8ZDR6 yersinia pe
30	133.5	5.5	1233	16	Q9ZSD7	Q9ZSD7 rhizobium m
31	133.5	5.5	1214	16	Q98M03	Q98M03 rhizobium 1
32	132.5	5.5	654	16	Q9A4U7	Q9A4U7 caulobacter
33	132	5.5	6077	12	Q9LH85	Q9LH85 white spot
34	131.5	5.5	553	2	Q9L2M2	Q9L2M2 rhodobacter
35	131.5	5.5	1248	16	Q8U101	Q8U101 agrobacteri
36	131.5	5.5	1292	10	Q9LPM4	Q9LPM4 arabidopsis
37	131.5	5.5	1309	10	Q9CAP4	Q9CAP4 arabidopsis
38	131.5	5.5	3501	16	Q8Y106	Q8Y106 ralsstonia s
39	131	5.4	409	16	Q56025	Q56025 salmoneilla
40	130	5.4	1953	16	Q98HU2	Q98HU2 rhizobium 1
41	129.5	5.4	1794	9	Q9TIA7	Q9TIA7 bacterioph
42	129.5	5.4	3072	2	Q93JN5	Q93JN5 streptococ
43	129	5.4	637	16	Q9AXK8	Q9AXK8 caulobacter
44	128	5.3	546	2	Q93G07	Q93G07 lactobacill
45	128	5.3	2432	12	Q8QZQ6	Q8QZQ6 chilo tride

ALIGNMENTS

RESULT 1
Q9DP34 PRELIMINARY; PRT; 486 AA.
ID Q9DP34
AC Q9JP34;
DT 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE Hypothetical 50.7 kDa protein.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=98422476; PubMed=9748456;
RA Charkoweki A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
RA Collier A.;
RT "The Pseudomonas syringae pv. tomato HrpW protein has domains similar
RT to harpins and peptidate lyases and can elicit the plant hypersensitive
RT response and bind to peptate.";
RL J. Bacteriol. 180:5211-5217(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=20243785; PubMed=10781092;
RA Alfano J.R., Charkoweki A.O., Deng W.L., Badel J.L.,
RA Petnicki-Ocwieja T., van Dijk K., Collier A.;
RT "The Pseudomonas syringae Hrp pathogenicity island has a tripartite
RT mosaic structure composed of a cluster of type III secretion genes
RT bounded by exchangeable effector and conserved effector loci that
RT contribute to parasitic fitness and pathogenicity in plants.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Ramos A.R., Rehm A.H., Collier A.R.;
RT "Pseudomonas syringae pv. tomato DC3000 hrpL through hrpU";
RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF232004; AAF71504.1; --
KW Hypothetical protein.
SQ SEQUENCE 486 AA; 50709 MW; 91CFBC03397F8AD1 CRC64;
Query Match 77.7%; Score 1869.5; DB 2; Length 486;
Best Local Similarity 78.4%; Pred. No. 1.5e-98;
Matches 381; Conservative 33; Mismatches 71; Indels 1; Gaps 1;
Qy 1 MHINSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLOAI TDYLKHHVFAAHRFS 60
Db 1 MYINRSISSOSSIGTSEFHSQA--SVASSSVRLSEABQTKLNDITHYLTGDFVFAAHLRP 58
Qy 61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
Db 59 VNDINVDQDVVHAHVEIGKII SARALRLDDEGETALSIGETFAKAEKFDRLMAASAGAL 118
Qy 121 ENTFFAAASVLQYMQPAINKGDMWLPKPLTPLISGALSGAMDQVGTMMDRARGDLHY 180
Db 119 RATPFMAASLQYMQPAINKGDMWLPKPLTPLISGALSGAMDQVGTMMDRATGDLHY 178
Qy 181 LSTSPDKLHDAMAVSVKHSRSPALGRQVVDMGIAVOTFSALNVVTVLAPALASRPSVOGA 240
Db 179 LSTAPEKLDHMAASVVRHHPGVNRQAADIGIAVQVTVARNALRTVLAPALASRPAVOGA 238
Qy 241 VDFGVSTAGGLVANAGFGRMLSVQSRDQLRGGAFLVGMKDEKPKAALSEETDMLDAYKA 300
Db 239 VDISVSAAGSLVANAGFSEMHVTVQARDHLRGGAFLVGIKDKQPKADLSEETDMLDAYRA 298
Qy 301 IKSASYSGAALNAGKRMAGLPDVAITDGLKAVRSLSVTSITKNGLALAGGAGVSKLOK 360
Db 299 IKSASYSGAALNAGKRVAGLPDVAITDGLKAVRSLSVTSITKNGLALAGGAGVSKLOE 358
Qy 361 MATKNIITDSATKAAVSQSLNVLGVSGVFAGMTTAGLATDPAVKKAESFIQDKVKSTASST 420
Db 359 MATKNIITHPATKAAVSQSLTNLAGSAPVFSATTTAAVATDPAVKKAESFLODTVKTTSVNG 418
Qy 421 TSYVADQTVKLAKTVKMSGEAISSTGASLRSTVNNLRHSAPADIEEGGISAFSRS-- 478
Db 419 TSALADTKVLAKAGIDASAERIATTTGASLRDT---LRRTVREPDIIEGGVAAAGTOSPV 475
Qy 479 ETPQLER 486
Db 476 AVPEAMR 483
RESULT 2
Q8RP03 PRELIMINARY; PRT; 484 AA.
AC Q8RP03;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Type III effector HopPtoAlpma.
OS Pseudomonas syringae pv. maculicola.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=59511;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ES4326;
RX MEDLINE=21862332; PubMed=11872842;
RA Guttman D.S., Vinatzer B.A., Sarkar S.F., Ranall M.V., Kettler G.,
RA Greenberg J.T.;
RT "A Functional Screen for the Type III (Hrp) Secretome of the Plant
Pathogen Pseudomonas syringae";
RL Science 295:1722-1726(2002).
DR EMBL; AF458051; AAL84253.1; --
SQ SEQUENCE 484 AA; 50520 MW; 4DB012DD5688163E CRC64;
Query Match 66.1%; Score 1591.5; DB 2; Length 484;
Best Local Similarity 67.6%; Pred. No. 9.7e-83;
Matches 330; Conservative 51; Mismatches 100; Indels 7; Gaps 3;
Qy 16 ESFRTASDASLASSSVRSVSTTSCRDLOA-----ITD 47

Qy 1 MHINSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLOAI TDYLKHHVFAAHRFS 60
Db 1 MYINRSISSOSSIGTSEFHSQA--SVASSSVRLSEABQTKLNDITHYLTGDFVFAAHLRP 58
Qy 61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
Db 59 VNDINVDQDVVHAHVEIGKII SARALRLDDEGETALSIGETFAKAEKFDRLMAASAGAL 118
Qy 121 ENTFFAAASVLQYMQPAINKGDMWLPKPLTPLISGALSGAMDQVGTMMDRARGDLHY 180
Db 119 RATPFMAASLQYMQPAINKGDMWLPKPLTPLISGALSGAMDQVGTMMDRATGDLHY 178
Qy 181 LSTSPDKLHDAMAVSVKHSRSPALGRQVVDMGIAVOTFSALNVVTVLAPALASRPSVOGA 240
Db 179 LSTAPEKLDHMAASVVRHHPGVNRQAADIGIAVQVTVARNALRTVLAPALASRPAVOGA 238
Qy 241 VDFGVSTAGGLVANAGFGRMLSVQSRDQLRGGAFLVGMKDEKPKAALSEETDMLDAYKA 300
Db 239 VDISVSAAGSLVANAGFSEMHVTVQARDHLRGGAFLVGIKDKQPKADLSEETDMLDAYRA 298
Qy 301 IKSASYSGAALNAGKRMAGLPDVAITDGLKAVRSLSVTSITKNGLALAGGAGVSKLOK 360
Db 299 IKSASYSGAALNAGKRVAGLPDVAITDGLKAVRSLSVTSITKNGLALAGGAGVSKLOE 358
Qy 361 MATKNIITDSATKAAVSQSLNVLGVSGVFAGMTTAGLATDPAVKKAESFIQDKVKSTASST 420
Db 359 MATKNIITHPATKAAVSQSLTNLAGSAPVFSATTTAAVATDPAVKKAESFLODTVKTTSVNG 418
Qy 421 TSYVADQTVKLAKTVKMSGEAISSTGASLRSTVNNLRHSAPADIEEGGISAFSRS-- 478
Db 419 TSALADTKVLAKAGIDASAERIATTTGASLRDT---LRRTVREPDIIEGGVAAAGTOSPV 475
Qy 479 ETPQLER 486
Db 476 AVPEAMR 483
RESULT 3
Q8XQF0 PRELIMINARY; PRT; 518 AA.
AC Q8XQF0;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Putative transmembrane protein.
GN RSP1277 OR RS05322.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502(2002).
DR EMBL; AL646083; CAD18428.1; --
KW Plasmid; Complete proteome.
SQ SEQUENCE 518 AA; 53581 MW; 82D486715F0F60EE CRC64;
Query Match 19.1%; Score 460; DB 16; Length 518;
Best Local Similarity 32.0%; Pred. No. 2.3e-18;
Matches 139; Conservative 63; Mismatches 159; Indels 74; Gaps 12;
Qy 16 ESFRTASDASLASSSVRSVSTTSCRDLOA-----ITD 47

RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurolo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmieu K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58";
RL Science 294:2323-2328 (2001).
DR EMBL; AE009328; AAL44779.1; ALT_INIT.
DR EMBL; AE008285; AAK89449.1; --
KW Complete proteome.
SQ SEQUENCE 1009 AA; 100468 MW; 14D8428308E5EE CRC64;

Query Match 6.2%; Score 149; DB 16; Length 1009;
Best Local Similarity 22.4%; Pred. No. 2.9;
Matches 127; Conservative 77; Mismatches 228; Indels 136; Gaps 25;

Qy 7 AQOPGVAMESFRTASDASLASSSVSVSTTSCRDLOAI-TDYLKH---HVFAAHRFSVI 62
Db 118 ASQIDAMSTDOIKALNSQVAGLSSAQVATLSSDELALFTTDEIKSISANAIGLSAAAI 177
Qy 63 GSPDERDAALAHNEQIDALVETFRANRLYSEGETPATIAETFAKAEKF-----DRLATTA 116
Db 178 AGLSTDNAALTKSQIAMSTQFNALTSGLATFSADEVKAINSKILLAGLDVTKLSTGN 237
Qy 117 SSAPENTPFAASVLYQMPAINKGMDLPLKPLTPLISGALSGAMDQVGTQMDRARG 176
Db 238 IAAASKAQSALSTQTPAAMSTDQ-----IKALTSQVAGLSSA--QVATLSSD--- 284
Qy 177 DLHYLSTSPDKLHDAMAVSVKHSFALGRQVDMGIA-----VQ 215
Db 285 ELALFSTDEIKAIKANAVA--GLSAAALAAALTDDNAALTKTQIAGLSLTQNALTSANLA 343
Qy 216 TFSALNVVTVLAPALASRPSVQGVDFGVSTAGLVA-----NAGFGDRLSVQSRDQL 270
Db 344 TFSA-DEIKAIKTRALA-----GLDVKLSTGNIALTKTOAASLSTQFAMSTDOI 395
Qy 271 RG--GAFVLMKDKPEKKAALSEE-----TDWLDAYKAIKSAYSGAALNA----- 313
Db 396 KALTSEQVAGLSSAQVATLSSDELALFTDEISAFSANAVAGLSAALAAALTGNATALT 455
Qy 314 KRRNAGL---PLDVAIDGLKAVRSVLSATSLTKNGLAGGAGVAGVSKLOKATKNTD 370
Db 456 KTOIAGLSTQNALTSGLSIATFSADEKAIKTR--ALAG-----LDVTKLSTGNVA-AL 507
Qy 371 TKAASVQLSNLVGSGVGFAGWT-----AGLATDPAVKKAES-----FIQDKVKS 415
Db 508 SKAQVSALSTT-----QFAMSTDOI KALTSEQVAGLSSAQVATLSSDELALFTDEIKA 562
Qy 416 -----TASSTTSYVADQTVKLAKT-VKDMGGEAISSTGASLRSTVNNLRHRSAP 464
Db 563 IGANAVAGLSAAALAAALTDDNASALTKTQIAGMSSTQINAL-----TSANLATFSAD 616
Qy 465 -----ADIEGGISAFSRSE 479
Db 617 KAITTKALGGLDVTKLSTGNIAALTAKQ 644

RESULT 6
Q92K98
ID Q92K98 PRELIMINARY; PRT; 2089 AA.
AC Q92K98;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical transmembrane protein SMC00190.
GN R01816 OR SMC00190.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Fumelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
DR EMBL; AL591788; CAC46395.1; --
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR003880; Ppantex_attach.
DR Pfam; PF00015; MCPsignal; 1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2089 AA; 222522 MW; 1D6334385A31ADAD CRC64;

Query Match 6.1%; Score 146; DB 16; Length 2089;
Best Local Similarity 18.9%; Pred. No. 12;
Matches 106; Conservative 84; Mismatches 208; Indels 162; Gaps 17;

Qy 3 INQSAQPPGVAMESFRTASDASLASSSVSVSTTSCRDLOAI-TDYLKHVFHFAHRFSVI 62
Db 188 LRSAAARSMAEVAM---RLAEPETNAADRVMTVGQAVREVSAMNEGERTIARATELEAL 244
Qy 63 GSPD-----ERDAALAHNEQIDALV-----E 83
Db 245 VHSVSALESYSSENELRVTLVQELGLEREALIGHSDRIETAIAGATKLDLETASE 304
Qy 84 TRANRLYSEGETPATIAETFAKAEKFDRLATTASSAPENTPFAASVLYQMPAINKGDM 143
Db 305 DIASRIASVGEAFASLIDTRAAA---LTDKSDHALEN----- 338
Qy 144 LATPLKPLT-PLISG-----ALSGAMDQVGTQMDRARGDLHYLSTSP 185
Db 339 LSTMLTTRTALLSGLTTAGVALSNEFDARLDALSDNLTORGEQLLSQFETRASTLDANT 398
Qy 186 DKLHDAMAVSVKHSFALGRQVDMGIAVQTFSAVNTVTLAPALASRPSVQGVDFGV 245
Db 399 EKLNAALNERAQNLNETLIARTDLNESLR-----IGQQAISGLDDVL 442
Qy 246 STAGLVANAGFGDRLMSVQSRD-----QLRGAP-----VLGMDKKEPKA 286
Db 443 SSLNSALDEKSGASPROSLKSSADDAIMDLRGGFPEKLTQTVGLASAFDERFHEFAS 502
Qy 287 ALSEETDMLDAYKAIKSAYSGAALNAGKEMAGLPLDVAIDGLKAVRSVLSATSLTKNGL 346
Db 503 AFDKRASQLDT-KLMSLHRINETVSGGSEAI GGALDSSVD---KINSALSEQSLT---L 555
Qy 347 ALAGGVAGVSKLOKATKNTDTSATKAASVQLSNLVGS-----VGVPAGWTTA 394
Db 556 ATALG-----ATQDFIETIGTSRSTSELSLIGNAHNRIESVLSDKTSGLMGALTE 605
Qy 395 -----GLATDP-----AVKKAESFIQDKVKKSTTSYVADQTVKLAKTVMKSGEA 442
Db 606 AQERIENGFGORADALANALTTGERSLTDGLDRTSFAFIEGLQSAHARIETGLTGSTDEI 665
Qy 443 ISSTGASLRSTVNNLRHSA 462
Db 666 TSAIASQHRDLNLTLSRTA 685

RESULT 7
O85472
ID O85472 PRELIMINARY; PRT; 2055 AA.
AC O85472;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Extracellular matrix binding protein (Fragment).
GN ENB.

OS Abiotrophia defectiva.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Aerococcaceae; Abiotrophia.
 ON NCBI_TaxID=46125;
 RP SEQUENCE FROM N.A.
 RC STRAIN=NVS-47;
 RX MEDLINE=9081722; PubMed=9864195;
 RA Mangenell R., van de Rijn I.;
 RT "Characterization of emb, a gene encoding the major adhesin of
 RT Streptococcus defectivus."
 RL Infect. Immun. 67:50-56(1999).
 DR EMBL; AF067776; AAD03320.1; -
 DR TIGRFBMS; TIGR01168; YSIRK_signal; 1.
 FT NON TER 2055 2055
 SQ SEQUENCE 2055 AA; 215642 MW; 9699C11DDE93E2FD CRC64;

Query Match 6.0%; Score 143.5; DB 2; Length 2055;
 Best Local Similarity 22.2%; Pred. No. 16;
 Matches 123; Conservative 78; Mismatches 249; Indels 105; Gaps 23;

QY 3 INGSQOOP-----PGVAMESFRTASDASIA--SSVSRSVSTTSCRD--LOAI-----TDY 48
 DB 726 INESQOPDLTREKQAFMOQVTRARDAAMAKVASAANQAVTSARDOGLMAVNNLTPA 785
 QY 49 LKHVFPAHFRFSVSGSPDERDALAHN-----EQIDAL-----VETRAVRLYSEGETPA 97
 DB 786 AKYPEALGH---VRQADAKRQAIRDNANLTAEQADALRQVDAQRAEAIAINQNTNA 842
 QY 98 TIATFPKAEKF-----DLATTAASAFENTPFAAASVQVQPIINKGWL-----A 145
 DB 843 TLKADSDGVKAIINDINPQPSKPAQAQLEQV--AAAK-----ROAINNNQLTDEKA 895
 QY 146 TPLKPLPLISGA---LSGAMQVGTGMDBARGDLHYLSTSPDKLHDAMAVSKRSPA 202
 DB 896 QALIQVQALANAKTQVQAANDNNGVNOAKTA--CTTAIINNIPGCTOKAQAIIEAEQ 954
 QY 203 LGRQVDMGIAVQTFSAIINVRYTLAPALASRPVQ-----GAVDFGVSTAGSLVA 253
 DB 955 AKRLEQGRNDLTTEBRNNMLADLTAKAQAKADVANOANNTGVAGAKKNGVQAQIQINP 1014
 QY 254 NA-----GFGDRMLSVSRDQLRGGAFLGKDKKEPKALSEETDWT--DAYKA 300
 DB 1015 TAVVPPARNAIDQAPARDKEAEFQANTKL-----TDEKAAAIKKYQDAPARDAKAA 1065
 QY 301 IKASVSG--AAINAGKPMAGPLDVATGKLAV--RSIVSATSLTNGL-----A 347
 DB 1066 IDRAGSNGDVNNNAVNOQK-----AAIQAIKALDSDQPAKDTATRAAIQNAADAKAA 1117
 QY 348 LAGGAGVSKLQKMATKNTDTSATKA--AVSQLSNLVGSVGFAGMTTAGLATDPAYKAAE 406
 DB 1118 ITANNALTOBEKKAALIQVEBEAKAQAAVADASRSKADVRAKQGIQKISDVRAVPPK 1177
 QY 407 SFIDOKYKSTASTSTSVADQTVKLATVQKMSGEAISSTGASIRSTVNNLRHSAPBAD 466
 DB 1178 LMAIAVDOQATDKKAVIINDTT--LTQBEKAAIRKVEDEAKARQAINDATSNADVAAK 1236
 QY 467 IEESGISAFSSETP 481
 DB 1237 QAQGTQAINNVPTP 1251

RESULT 8
 Q9HPA2 PRELIMINARY; PRT; 536 AA.
 AC Q9HPA2;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Hcr17 transducer.
 GN HTR17 OR VNG1733G.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;

OC Halobacteriaceae; Halobacterium.
 ON NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berguiet B., Pan M.,
 RA Snukia H.D., Laaky S.R., Baliga N.S., Thorsson V., Sbrigna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
 RA Leitnauber B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocke D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Lemminger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ehardt H., Lowe T.M., Liang P., Riley W., Hood L., Dasgupta S.,
 RT "Genome sequence of Halobacterium species NRC-1."
 RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL; AE005078; AAG19968.1; -
 DR HSSP; P02942; 1QJ7
 DR InterPro; IPR004089; Chmtaxis_trained.
 DR InterPro; IPR003660; HAMF.
 DR InterPro; IPR004090; Me_chemotaxis.
 DR Pfam; PF00015; MCPsignal; 1.
 DR PRINTS; PR00260; CHEMTRNSDUCR.
 DR SMART; SM00304; HAMF; 1.
 DR SMART; SM00283; MA; 1.
 KW Complete proteome.
 SQ SEQUENCE 536 AA; 56815 MW; 3B33D2FD8701BE04 CRC64;

Query Match 5.9%; Score 141; DB 17; Length 536;
 Best Local Similarity 23.2%; Pred. No. 3.3; Indels 120; Gaps 23;
 Matches 119; Conservative 78; Mismatches 197;

QY 12 GVAMESFRTASDASIASVSVSTTSCRDQAITYLKHVFAHFRFSVSGSPDERDAA 71
 DB 63 GPVMDAFST-----SLAVLIGVAIAISTGTISIEVD-LKTSNBARH---RRKDAEQAK 113
 QY 72 LANEQIDALVETRAVRLYSEGETPATIAETPAKAEKFDRLATTASAFENTPFAAASVL 131
 DB 114 RRADEQPAQOAEKAEKADREAEFTLAALQERAAERDQLSKASESVSVEKAS-- 171
 QY 132 QVQPIINKGDMTLATLKPPLI--SGALSGAMDQV-----TKMDRAR----- 175
 DB 172 -----GD-----LTVRANSSSDSDANTEVQGFENEMIAAFEEVARTREFAGD 214
 QY 176 -----GDLHYLSTSPDKLHDAMAVSVKRSFALGRQVDMGIAVQTFSAIINVRYTLAPA 230
 DB 215 VADNSNDIHASVDEIRASQVAESVQETISAGTDR-----HTRLADA 257
 QY 231 LASRPSVQAVDFGVSTAGGLVANAAGFGDRMLSVSRDQLRGAFVYLGKDKPEKALSE 290
 DB 258 TDEIQSLSGAVE--EVAAASQVAD-----VSDQADRRGQS-----DDAAAATVQ 302
 QY 291 EFTWLDAYKAIKS--ASVSGAALNAGKRMAGPLDVATD--GLKAVSISVATSLTKGLA 347
 DB 303 MTDIQSTETAEVSEANALAEAGEIRRVYDLINDI--TDQTHLALANASIEAARADGDG 361
 QY 348 LAGGAGVSKLQKMATKNTD-----SATKAASQLSNLVGSVGFAGMTTAGLA 397
 DB 362 FAYVAAEVEKLAE--ETNEATDEIETLVNLLSTQSVVDSEGMATVS--SGTEVENA 418
 QY 398 -----TDPAYKKAESFTQDKYKSTASTSTSVADQTVKLAKTVQKMSGEAISST 446
 DB 419 LTLALBEIGHVTD--VNEAAVQIIDAATDAQAA--ADTVSMIEIRIDIS--EQNSSSE 470
 QY 447 GASLRSTVNNLRHSAPBADIEEGGISAFSRSST 480
 DB 471 AASVSA--QOOTASLSDVTGVTTLVLRADT 501

RESULT 9
 Q8XSW6 PRELIMINARY; PRT; 642 AA.
 ID Q8XSW6;
 AC Q8XSW6;
 DT 01-MAR-2002 (Tremblrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Probable flagellar hook-associated protein 1.
 GN FLGK OR RSP0351 OR RS00754.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Plasmid megaplasmid
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RA MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arlat M., Billault A., Brotter P., Camus J.C., Cattolico L.,
 RA Chaudier M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gapin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
 RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 DR EMBL; AL646078; CADI7502.1; -
 DR InterPro; IPR001444; Flag_bb rod.
 DR Pfam; PF00460; flg_bb rod_1
 DR PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 642 AA; 65396 MW; 585D4CF4A6989313 CRC64;
 Query Match 5.8%; Score 140.5; DB 16; Length 642;
 Best Local Similarity 20.9%; Pred. No. 4.5; Indels 167; Gaps 21;
 Matches 119; Conservative 67; Mismatches 216;
 QY 12 GVAMESFRITASLASLA-----SSSVRSVSTTSCRDQLAITYLKHVFAAHRFSVIGSPDE 67
 Db 108 GSALTFFENAD-SLASNPSTTSARQVLSAATLTQT-----RFLAISG--- 150
 QY 68 RDAALAHNEQDALVETRANRLYSGETPATIAETFAKAEKFDRLATTASSAFENTPPAA 127
 Db 151 QMASLSN--QVNTQVQTVNSVNSTQIOALNDQIAKAE-----AASGG----- 193
 QY 128 ASVLQYMPAINKGDWLTATPLKPLTSLGALSGAMD-----QVGTKMDRARGDLHYLS 182
 Db 194 -----QFANDLRQRDLQVLTNQSIVASVQTSQDQYNYVNGQA-LVQNGSQQL 245
 QY 183 TSPDKLHDAMAVSVKRRSPALGRQVDMGIAVQTFPSALNVVR-TVLAPALASRPSVQGV 241
 Db 246 TTVASQYDPTQLSVGKSPAGVTITDDSQLGGGLGGLMQFRONTLIPQNSLGLRAAV 305
 QY 242 DRGV-----STAGGLVA-----NAGFGDRMLSVQSRD-----Q 269
 Db 306 SADVNTQNKEGMDLNGKLGTDLTFSTAGSPVSAASSNTGTGALTATITNANAGQGYQVK 365
 QY 270 LRGGAFVLG-MKDKEPKAALSEETDMLDA-----VDFGVSTAGLVANAGFGDRMLSVQSRDQLRGAF 297
 Db 366 YSGGAYTVSHYPDGGAVTVSSWPTTVQGVTLNLTGSMASGDSFLVRPTVNAASTMQTLT 425
 QY 298 ---YKAIKASYSYGAALNAGKRWAGLP--LDVATDGLKAVRSIVSATSITKNGLAGALAG--- 350
 Db 426 SDYHAVA---SPVVNQGSSNNTGTSVASIGVDSTYAGSPLASAVSLTYSGSLSGFPFG 483
 QY 351 -----GYAGVSKLQKVA-----TKNITDSATKA 373
 Db 484 SVTVTVNGTATTYSGTAPYTOGATYSFNGIQMSLTGTTPAANDTFTVSANTANSTDGHNAS 543
 QY 374 AVSOLSNLV-----GSGVGFAGTWTAGLATDPAVKAESFIQDKVKSTASTSTTSYVADQTV 429
 Db 544 APAQLRNATVLDNGTTSISSGNSNLVTVQVIGIQASQASANLITSQKALLASSTSQQSVSGV 603
 QY 430 KL-AKTVKDMSGEAISSTGASLRSTVNNL 457
 Db 604 NLDDEAMNLMKYQQAQVQASAKVMQNTANSL 632

RESULT 10
 Q931N9 PRELIMINARY; PRT; 1983 AA.
 ID AC Q931N9;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Mrp protein.
 GN MRP OR SAV1758.
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Staphylococcus.
 OX NCBI_TaxID=158878;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21311952; PubMed=11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizukami Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kihara S., Goto S., Yabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 aureus";
 RL Lancet 357:1225-1240(2001).
 DR EMBL; AF003363; BAB57920.1; -
 DR TIGRFAMs; TIGR01168; YSIRK_signal; 1.
 KW Complete proteome.
 SQ SEQUENCE 1983 AA; 215786 MW; 573C6E6DFFC7FBF CRC64;
 Query Match 5.8%; Score 140.5; DB 16; Length 1983;
 Best Local Similarity 19.3%; Pred. No. 23;
 Matches 105; Conservative 94; Mismatches 203; Indels 143; Gaps 21;
 QY 20 TADASL-----ASSVRSVSTTSCRDQLAITYLKHVFAAHRFSVIGSPDERDAALAHN 75
 Db 930 TAANTNLNANTNADVEQVKNATQIGQIATPATK-----VKTAKNA----- 972
 QY 76 EQDALVETRANRLYSGETPATIAETFAKAEKFDRLATTASSAFENTPPAAASVLQYMQ 135
 Db 973 --IDKSAETQHTNIFNND--ATLEEQAAQQLLDQAVATAKQIN-----AADTQGEVA 1023
 QY 136 PAIKNGDWLTATPLKPLTSLGALSGAMDQVGTMDRARGDLHYLSSTSPDKLHDAMAVS 195
 Db 1024 QAKDQGTQNTVVIQPAQVKTARDNA-----VNDKAREAITNINATPGATREEKQEA 1075
 QY 196 VKRHPALGRQVDMGIAVQTFPSALNVVR-----VLAPALASRPSVQGA----- 240
 Db 1076 INRVNTLKNRALTDTIGV-TSTTAMVNSIRDDAVNQIGAVQPHVTKQTATGVNLNDLATAK 1134
 QY 241 -----VDFGVSTAGLVANAGFGDRMLSVQSRDQLRGAF----- 275
 Db 1135 KOEINQNTNATTEEKQVALNQVQDELATA---INNINQADTNAEVDQAOQLGTKAINAIQ 1191
 QY 276 -----VLGMKDKPEKALSEETDMLDAYKAIKASYSYGAALNAGKRW----- 318
 Db 1192 PNIVKPEAALAQINQHYNAKLAENATPDATNDEKNAIN--TLNQDROQAIESIKQANT 1249
 QY 319 GLPLD-----VATDGLKAVRSIVSATSITKNGLAGALAGVAGVSKLQKVA--TKNITDSATK 372
 Db 1250 NAEVDQAAQTVANNIDAVQVDVVKQQAARDKIT-----AEVAKRIEAVKQTPNATDEKQ 1304
 QY 373 AAVSOLSNLVSGVGFAGTWTAGLATDPAVKA-ESFIQDKVKSTASTSTSYV-----ADQ 427
 Db 1305 AAVNQINQL-----KQAINQINQNTNDQVDTTNNQAVNAIDNVEAEV 1348
 QY 428 TVK-----LAKTVKMSGEAISSTGASLRSTVNNLHRSAPEADIEEGGISAPSRSETP 481
 Db 1349 VIKPKALADIEKAVKEKQQQIDN-----SLDSTDNKEVASQALAKEKEKALAAIDQATN 1404
 QY 482 FQLRR 486

Db 1405 SQVNO 1409

RESULT 11

O99TB0 PRELIMINARY; PRT; 2186 AA.

AC O99TB0; 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein SA1577.

GN SA1577.
 OS Staphylococcus aureus (strain N315).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Staphylococcus.

OX NCBI_TaxId=158879;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=21311952; PubMed=11418146;

RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 Cui L., Oguchi A., Aoki K.I., Nagai Y., Iian J., Ito T., Kanamori M.,
 Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
 Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
 Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
 Yamashita N., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
 Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 aureus".
 RL Lancet 357:1225-1240(2001).

DR EMBL; AP003135; BAB42845.1; -;
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR TIGRPFAM; TIGR01167; LPXTG_anchor; 1.
 DR TIGRPFAM; TIGR01168; YSIRK_signal; 1.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KM Hypothetical protein, Complete proteome.
 SQ SEQUENCE 2186 AA; 23848 MW; 0CC84636F1868A4F CRC64;

Query Match 5.8%; Score 140.5; DB 16; Length 2186;
 Best Local Similarity 19.3%; Pred. No. 27;
 Matches 105; Conservative 94; Mismatches 203; Indels 143; Gaps 21;

QY 20 TASDASL-----ASSVSRSVSTTSCRDLOAITDYKHHVFAHRSVIGSPDERDALAHN 75
 DB 930 TAANTNINANTNADVQVKTNAIQIGIATPATK-----VKTDAKNA----- 972
 QY 76 EQIDALVETRANRLYSGETPATIAETFAKAEKEDRLATTTASSAFENTPPPAASVLOQM 135
 DB 973 --IDKSAETHNTFFNNND--ATLEEQAAQQLDQAVATRAKONIN-----ADTNGEVA 1023
 QY 136 PAINKGDMATPLKPLTPLISGALSAMDQVGTGMDPARGDLHYLSTSPDKLHDMAVS 195
 DB 1024 QAKDQCGQNVIVLOPAQOVKTDAANA-----VNDKAPRAITNINATPGATTEBEQEA 1075
 QY 196 VKRSPALGQVVDWGIAVOTFSALNVRT-----VLAPALASRPSVOGA----- 240
 DB 1076 INRVNTLKNALTDIGV--TSTTAMVNSIRDDAVNIGAVGPHVTKQTATGVINDLATAK 1134
 QY 241 -----VDGVSITAGGLVANAGRGMRMSVSGRDLRGAF----- 275
 DB 1135 KOEINONTNATTEBEKQVALNOVDDELATA--INNINQADTNAEVDAQOQGTAINAIQ 1191
 QY 276 -----VLGMKDEPKPAALSEETDMLDAYKAIKSASYSAGALNAGKRNA----- 318
 DB 1192 PNTVKKRAALAIQNHNAKLAEINATPDPATNDEKMAAIN--TLNORQQAIBSITKQANT 1249
 QY 319 GLPLD---VATDGLKAVRSIVSATSLTKGALAGGAVAGSKLQKMA--TKNITDSATK 372
 DB 1250 NAEVDQATVAENNIDAVQVVDVKKQARDKIT-----AEVAKIKIENVKQPNATDEBEQ 1304
 QY 373 AAVSQSLNVIGSVGVFAGMTTATGATDPAYKKA--ESFIDQVKSITASTTSYV---ADQ 427
 DB 1305 AAVNQINQL-----KQALINQINQNTNQDVDTTNOAVNAIDNVEAEV 1348

QY 428 TVK-----LAKTVKMSGEAISSTASLSRSTVNNI.RHRSAPADIEEGISAPSSSETP 481
 DB 1349 VIKRAIADIEKAVKEKQOQIDN-----SLDSTNEKEVASQALAKEKEKALAIIDQACTN 1404
 QY 482 FOJLR 486
 DB 1405 SQVNO 1409

RESULT 12

O9FC05 PRELIMINARY; PRT; 503 AA.

AC O9FC05; 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Flagellin (Fragment).

GN FLIC.
 OS Pseudomonas fragi.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxId=296;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=NCIMB8542;
 RA Beilingham N.F., Morgan J.A.W., Saunders J.R., Winstanley C.;
 RT "Flagellin gene sequence phylogeny in the genus Pseudomonas".
 RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL; AJ297534; CAC03723.1; -;
 DR InterPro; IPR001492; FlagellinN.
 DR InterPro; IPR001029; Flagellin_C.
 DR Pfam; PF00700; Flagellin_C; 1.
 DR Pfam; PF00669; Flagellin_N; 1.
 DR PRINTS; PR00207; FLAGELLIN.
 DR Prodom; PD000316; Flagellin_C; 2.
 FT NON_TER 1
 FT NON_TER 503

SQ SEQUENCE 503 AA; 51065 MW; AE42EFB521BF7C1 CRC64;

Query Match 5.8%; Score 140; DB 2; Length 503;
 Best Local Similarity 24.3%; Pred. No. 3.4;
 Matches 99; Conservative 43; Mismatches 135; Indels 130; Gaps 19;

QY 147 PLKPLPLISGA-----LSGAMDQVGTGMDPARGDLHYLSTSPDKLHDMAVSVK 197
 DB 10 PVKGAVTIANANNGSSIAQTAEAGM--QESTNITLQIR--ELALQSANGLKSDPRASLQ 67
 QY 198 RHSPALG-----ROYVDWGIAVOTF----- 217
 DB 68 EPTAKVGELTRISSTTFGSRNLLDSFGQSFQVGADANQTISFGMSDISATGLKSGY 127
 QY 218 --SALNVRTVLAPALASRPSVOGAVDFGVSTYGLVYANAGFDRMLSVSROQLRGAF 275
 DB 128 EASAAGVSTLSANV-----VEGANDATFKATGAAPFAVDDQTLITNGT----- 174
 QY 276 VLGMKDE-----PKAALSEETDMLDAYKAIKSASYSGA 309
 DB 175 IALAKSKIADAVABINQKTSKGTATASADATGTCTLTLSASDFRAVGSAAADAGFVAA 224
 QY 310 ALNAGKRMAGLPDVAITDGLK--AVRSIV--SATSLTK--NGLALAGGAVAGSKLQKMATK 364
 DB 235 ATPA--KNLGRPVEIQNGVKVITIAAGSLEMAAAITTKANTDSKGTGVNAGVKGRVLTS 293
 QY 365 -----NITDSATKAAVSQSLNVIGSVGVFAGMTTATGATDPAY-----KKASFT 409
 DB 294 EKOQAINLADSTGTGPGSLSKL-----GLTAGSTOKLTNDTSVSPNGVVKPKGDS-- 347
 QY 410 QDKVKTASS-----TTSYVADQTVLAKTVKMSGEAISSTASLSR 451
 DB 348 MDTIVSINSASTGTATASKADNTLKLFT--KD-----ITTAGGSCR 388

RESULT 13

064046 ID O64046 PRELIMINARY; PRT; 2285 AA.
AC O64046;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative transglycosylase.
GN YOMI.
OS Bacteriophage SPBc2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=66797;
RN [1]
RP SEQUENCE FROM N.A.
RA Lazarevic V., Dusterhoef A., Soldo B., Hilbert H., Maue L.,
RA Karamata D.;
RA "The complete nucleotide sequence of the Bacillus subtilis spbetac2
RT prophage";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF020713; AAC13005.1; -;
DR HSP; P00718; 153L.
DR InterPro; IPR002886; Peptidase M37.
DR InterPro; IPR000189; SLT domain.
DR Pfam; PF01551; Peptidase_M37; 1.
DR Pfam; PF01464; SLT; 1.
DR PROSITE; PS00922; TRANSGLYCOSYLASE; 1.
SQ SEQUENCE 2285 AA; 252310 MW; FF602C227754B357 CRC64;
Query Match 5.8%; Score 140; DB 9; Length 2285;
Best Local Similarity 20.5%; Pred. No. 30;
Matches 128; Conservative 80; Mismatches 224; Indels 192; Gaps 26;
QY 15 MESPTASDASLASSSVRSVSTTSCRDLOATDYKHHVFAHRFSV-IGSPDERDAALA 73
DB 251 IELYQRAQVNVQNLNTRYGSSMGSSNRQAVQDYLN----AVNSLNVSTGSSNNIRSQS 306
QY 74 HNEQIDAL---VETANRLYSGEPTATAEFF-----AKA 106
DB 307 LNNQPRELASNAQTAANQASSFG---AELTQFKMSYTLISGSLFYGAISGLKEWVQA 363
QY 107 EKFDRLATTASSAFENTPPAAASVLQYMQPAINKGDWLATPLKPIITPLISGALGAMQV 166
DB 364 IEIDTLTNIIRVMNEPDYKYNELL---QESIDLGLTSLNKITDILQMTGDFRGMGFDES 420
QY 167 GYKMDRARGDLHYLS-TSPDKLHDAMVSVKRRHPALGRQV-----VDMGLAVQTF 217
DB 421 ELSTLTKTQAQLQNVSDLPDPTVNTLTAAELNPNIAANDSISADKLEVDNRYAVTTL 480
QY 218 SALNVVTVLAPALASRPSVOGAVDFGVSTA-----GGLVAN-----AGFGDRMLS 263
DB 481 DLANSIRK--AGSTASTFGVELNDLIGYTATASTTRESGNVGNLSKTIPIRIGNQSS 538
QY 264 VOSRDQLRGCAFVLGMKKEPKAALSE---ETDWL-DAYKAIKASYSQ----- 308
DB 539 IKALEQIGISVKTAGEAKSASDLISEVAGKWDTLSDAQKQNTSIGVAGIYQLSRFNAM 598
QY 309 ----AALNAGK----- 315
DB 599 NNFSTAQNAAKTAANSTGSAWSEQQKYADSLQARVKNLQNNFTFAIAASDAFISDGLIE 658
QY 316 --RMAGLPDLDVADGKLAV---RSLVSATS-----LTKNGLALAG-YAGVSKL--QKMA 362
DB 659 FTQAAGSLNASTGVKISVGVFPPLPAVAVSTATLLSKNTRILASLLIUTRAMGOETLA 718
QY 363 TKNITDSATKAASVQ-----LSNLVGSVGVFAGWTAGLATDPA-VKKA----- 405
DB 719 TAGLEAGMTRAAVASRVLTALRGLLVSTLVGGAPALGWALESLSISPAEAKKAKDPE 778
QY 406 -----ESFQ-----DKVKSTASTTSYVADQ---TVKLAKT---VK 436
DB 779 QSQQTNVEAITTNKSDTKLQIQYKELQKVGSRSLTSDEEQEYLQVTOOLACTFPALVK 838
QY 437 --DMSGEAISSTGASLRSTVNNLR 458

DB 839 GYDSQGNAILTKNKELEKAIENTK 862
RESULT 14
O31976 ID O31976 PRELIMINARY; PRT; 2285 AA.
AC O31976;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE YOMI protein.
GN YOMI.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Arevedo V., Bextero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haiteh J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krohn S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue L., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche M., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzensegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis";
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z99115; CAB14053.1; -;
DR HSP; P00718; 153L.
DR InterPro; IPR002886; Peptidase_M37.
DR InterPro; IPR000189; SLT domain.
DR Pfam; PF01551; Peptidase_M37; 1.
DR Pfam; PF01464; SLT; 1.
DR PROSITE; PS00922; TRANSGLYCOSYLASE; 1.
KW Complete proteome.
SQ SEQUENCE 2285 AA; 252310 MW; FF602C227754B357 CRC64;
Query Match 5.8%; Score 140; DB 16; Length 2285;
Best Local Similarity 20.5%; Pred. No. 30;
Matches 128; Conservative 80; Mismatches 224; Indels 192; Gaps 26;
QY 15 MESPTASDASLASSSVRSVSTTSCRDLOATDYKHHVFAHRFSV-IGSPDERDAALA 73
DB 251 IELYQRAQVNVQNLNTRYGSSMGSSNRQAVQDYLN----AVNSLNVSTGSSNNIRSQS 306
DB 74 HNEQIDAL---VETANRLYSGEPTATAEFF-----AKA 106
DB 307 LNNQPRELASNAQTAANQASSFG---AELTQFKMSYTLISGSLFYGAISGLKEWVQA 363
QY 107 EKFDRLATTASSAFENTPPAAASVLQYMQPAINKGDWLATPLKPIITPLISGALGAMQV 166
DB 364 IEIDTLTNIIRVMNEPDYKYNELL---QESIDLGLTSLNKITDILQMTGDFRGMGFDES 420
QY 167 GYKMDRARGDLHYLS-TSPDKLHDAMVSVKRRHPALGRQV-----VDMGLAVQTF 217
DB 421 ELSTLTKTQAQLQNVSDLPDPTVNTLTAAELNPNIAANDSISADKLEVDNRYAVTTL 480
QY 218 SALNVVTVLAPALASRPSVOGAVDFGVSTA-----GGLVAN-----AGFGDRMLS 263
DB 481 DLANSIRK--AGSTASTFGVELNDLIGYTATASTTRESGNVGNLSKTIPIRIGNQSS 538
QY 264 VOSRDQLRGCAFVLGMKKEPKAALSE---ETDWL-DAYKAIKASYSQ----- 308
DB 539 IKALEQIGISVKTAGEAKSASDLISEVAGKWDTLSDAQKQNTSIGVAGIYQLSRFNAM 598
QY 309 ----AALNAGK----- 315
DB 599 NNFSTAQNAAKTAANSTGSAWSEQQKYADSLQARVKNLQNNFTFAIAASDAFISDGLIE 658
QY 316 --RMAGLPDLDVADGKLAV---RSLVSATS-----LTKNGLALAG-YAGVSKL--QKMA 362
DB 659 FTQAAGSLNASTGVKISVGVFPPLPAVAVSTATLLSKNTRILASLLIUTRAMGOETLA 718
QY 363 TKNITDSATKAASVQ-----LSNLVGSVGVFAGWTAGLATDPA-VKKA----- 405
DB 719 TAGLEAGMTRAAVASRVLTALRGLLVSTLVGGAPALGWALESLSISPAEAKKAKDPE 778
QY 406 -----ESFQ-----DKVKSTASTTSYVADQ---TVKLAKT---VK 436
DB 779 QSQQTNVEAITTNKSDTKLQIQYKELQKVGSRSLTSDEEQEYLQVTOOLACTFPALVK 838
QY 437 --DMSGEAISSTGASLRSTVNNLR 458


```

QY 74 HNEQIDAL---VETRANRLYSEGETPATIAETP-----AKA 106
DB 307 LNNQFRELNAQTAANQASSFG---AELTQTFKSMSTYLLISGLFYGAISGLKEMWSQA 363
QY 107 EKFDRLATTSSAFENPFPAAASVLYQMOPAINKGWMLATPLKRLTPLLISGALSGAMDQV 166
DB 364 IEDTLMTNIRRWNNEDDYKNNEL---QESIDIGDTLSNKITDILOMTGDFGMRGFDES 420
QY 167 GTTMMRPARGDLYLS-TSPDKLHDAMAVSKRSPALGROV-----VDMGIAVQTF 217
DB 421 ELSLTLTQAQVQLQVNSDLTDDTNTLTAAMLNFINIANDSISADKLANEDVNNVATTL 480
QY 218 SALNVAVTTLAPALASPPSYQGAADFVGVSTA-----GGLVAN-----AGFDRMLS 263
DB 481 DLANSIRK--AGSTASTFGEVLENDLIGYTTALASTTRESGNIYVNSLKTIFARIGNQSS 538
QY 264 VQSRDOLRGAFVLMGMDKPKALSE---ETDWL-DAYKAIKASISG-----308
DB 539 IKALBOIGISVKTAGGAKASADLISEVAGKMDTLSDAQKNTSIVAGIYQLSRFNAMM 598
QY 309 ----AALNACK-----315
DB 599 NFNISIAQNAKTAANSVGSAMSEQKYADSLQARVNTLQNNFTFPALASDAPISDGLIE 658
QY 316 --RMAGLPLDVAITDGLKAV---BSLVATS-----LTKNGLALAGG-YAGVSKL--QKMA 362
DB 659 FTOAGAGSLNASTGVKSVGFPLPLLAAVSTATLLSKNRTLASLILGTRAMGQETLA 718
QY 363 TKNITDSATKAAVSQ-----LSNIVGSVGVAGMTAGLADPA-VKKA-----405
DB 719 TAGLEAGMTAAVAASRVLKALRGLVSTLVGAFALGALLESLSFPAKAKXADFE 778
QY 406 -----ESFIQ-----DKVSTASTTSYVADQ---TVKLAKT---VK 436
DB 779 QSQGTNVEALTTNKDSTDKILQCKYKELQKXKESLSLTSDEQETLQYQLAQFPALVK 838
QY 437 --DMSGEAISSGTASLSTVNLR 458
DB 839 GYDSQGNAILKTNKELEKAIENTK 862

RESULT 15
Q93GT1 PRELIMINARY; PRT; 491 AA.
AC Q93GT1;
ID 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Flagellin (fragment).
GN FLA.
OS Campylobacter lari.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxId=201;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NCCTC12896;
RA Sekizuka T., Gondo T., Murayama O., Moore J.E., Matsuda M.;
RT "Genotypic and phenotypic characterization of flagellin from strains
RT of urease-positive thermophilic Campylobacter (UPTC).";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB073918; BAB71801.1; -
DR InterPro: IPR001492; Flagellin_N.
DR InterPro: IPR001029; Flagellin_C.
DR Pfam: PF00700; Flagellin_C; 1.
DR Pfam: PF00669; Flagellin_N; 1.
DR ProDom: PD000316; Flagellin_C; 2.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 491 AA; 50360 MW; 65080B0DB25FP3EE CRC64;

```

Query Match

5.8%; Score 139.5; DB 2; Length 491;

```

Best Local Similarity 23.8%; Pred. No. 3.5;
Matches 114; Conservative 66; Mismatches 183; Indels 117; Gaps 24;

QY 21 ASDASLA--SSSVSYSTTSCRDLOAITDYLKHHVFAARFVSIGSPDERDAL-----A 73
DB 23 ALDSSLARLSSSGRINSADADGALAIADSLKQ--ANTLGAQAINNANDANSMLQIADKA 80
QY 74 HNEQIDAL---VETRANRLYSEGETPATIA---ETPAKKEKFDRLATTA-----SSA 119
DB 81 MDRLKILDTIKKATQAADQGTAKTRAMIQGEINKLMELDNIANTTYNGKQLLSGS 140
QY 120 FENTPPAASVLYQMOPAINKGWMLATPLKRLTPLLISGALSGAMDQVGTMMBRAR----175
DB 141 FSNQOFQI-----GD---KANQTVNATIGSTNSAKIGQ--TRFETGARVTS 182
QY 176 GDLAHYLSTSPDKLHDAMAVSKRHS---PALGRQVDMG-IAVQT-FSALNVVTVLAPA 230
DB 183 GDIAITLKATYDGINDYKFGSVTVSTGVGGLALAEINKVSDQTVRASATVQTISTAA 242
QY 231 LASRPVQGAVDVGVSTAGLVANAGFGDRML--SVQSRDOLRG-----GAFVLMKD 281
DB 243 LTAGSTGEFTINGV-VIGKVFVKAGDKDQALVSAINAKDQTTGVEASIVDGKLVNSAD 301
QY 282 -----KEPKALSEE-----TDWLDAIKAIKSASYSGAALNAGRMAGLPL 322
DB 302 GRGIKLSGISGLDADQIAEENYGRLLTVKNDGSDIVSTTGAFSTGATVAG-----L 354
QY 323 DVATDGLKAVRSLVSATSLTKNGLALAGYAGVSKLQ-----KMATKNIT 367
DB 355 SEATVNLSEIKGEISA-----DIASMGFNMSKAOIGTKOSACVTTLQGMAYVNDIA 408
QY 368 DSATKAAVSQLSNLVGSVGVFAGMTAGL---ATDPAVKAESFLQDKVSTASTTSY 423
DB 409 DT-----AIALNDITRANIGATQNIISTINNISVTVQNVVQAASQIRD--VDPASSANY 462

Search completed: January 2, 2003, 15:20:39
Job time : 37.0308 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:14:20 ; Search time 9.50976 Seconds
(without alignments)
2124.025 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407
Sequence: 1 MHINGSAQCPGVAMESFRT.....EEGIGAFSRSSTPQLRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seque, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	138	5.7	1120	1 STFR_ECOLI	P76072 escherichia
2	137	5.7	540	1 CH60_LACHE	O68324 lactobacilli
3	130	5.4	500	1 FLUB_SALAE	P52615 salmonella
4	130	5.4	1025	1 SLAP_CAUCR	P35828 caulobacter
5	129	5.4	436	1 Y868_CHLMU	O99131 chlamydia m
6	128.5	5.3	2388	1 SPCP_RAT	O99988 rattus norv
7	127	5.3	397	1 NCCB_ALCXX	O45855 alcaligenes
8	123	5.1	544	1 CH60_BACHD	O50305 bacillus ha
9	121.5	5.0	1068	1 HIRP_MOUSE	O91KY5 mus musculu
10	120.5	5.0	1140	1 YM96_YEAST	O04893 saccharomyc
11	120	5.0	1306	1 MSB2_YEAST	P32334 saccharomyc
12	119.5	5.0	1608	1 HLYA_SBRMA	P15370 serratia ma
13	119	4.9	657	1 HS7E_CABEL	P11111 caenorhabdi
14	118.5	4.9	760	1 MS1E_MYCLE	O05564 mycobacteri
15	118.5	4.9	880	1 LYTD_BACSU	P39848 bacillus su
16	117	4.9	583	1 YP65_MYCTU	O50723 mycobacteri
17	117	4.9	682	1 GR78_YEAST	P16474 saccharomyc
18	116.5	4.8	570	1 FLIR_RHOSH	O51151 rhodobacter
19	116	4.8	763	1 HTR2_HALNI	O98911 halobacteri
20	116	4.8	764	1 HTR2_HALSA	P71410 halobacteri
21	116	4.8	1577	1 HLYA_PROMI	P14616 proteus mir
22	116	4.8	3591	1 PHAB_BORPE	P12255 bordetella
23	115.5	4.8	401	1 YOPB_YERPS	O06114 yersinia ps
24	115.5	4.8	1065	1 SED4_YEAST	P25315 saccharomyc
25	115	4.8	679	1 GR78_KULLA	P22010 kluyveromyc
26	115	4.8	1239	1 V120_EBY	P03189 epstein-bar
27	113.5	4.7	571	1 FLAI_CAMJE	P56963 campylobact
28	113	4.7	2541	1 TAL1_HUMAN	O94940 homo sapien
29	112.5	4.7	505	1 FLUB_SALTY	P52616 salmonella
30	112.5	4.7	2090	1 N214_HUMAN	P32658 homo sapien
31	112	4.7	439	1 Y579_CHLTR	O84593 chlamydia t
32	112	4.7	492	1 FLIC_SALRU	P06175 salmonella
33	112	4.7	1027	1 AF10_HUMAN	P55197 homo sapien

34	111.5	4.6	547	1 CH60_ALTRA	O9xau7 alteromonas
35	111.5	4.6	553	1 MCP2_ECOLI	P07017 escherichia
36	111	4.6	881	1 YJH8_YEAST	P47037 saccharomyc
37	110.5	4.6	543	1 CH60_MYCGE	P47632 mycoplasma
38	110.5	4.6	1419	1 ALAI_CANAL	O13368 candida alb
39	110	4.6	545	1 HTR5_HALNI	O48318 halobacteri
40	110	4.6	814	1 SLAI_BACAN	P49051 bacillus an
41	109.5	4.5	901	1 SOK1_YEAST	P40317 saccharomyc
42	109.5	4.5	2541	1 TAL1_HUMAN	P26039 mus musculu
43	109	4.5	1302	1 P3P_LACLC	P15292 lactococcus
44	109	4.5	2090	1 HRC1_MESAU	P51611 mesocricetu
45	108.5	4.5	465	1 SLAP_LACBR	O05044 lactobacilli

ALIGNMENTS

RESULT 1
STFR_ECOLI
ID STFR_ECOLI STANDARD; PRT; 1120 AA.
AC P76072; P7560;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Side tail fiber protein homolog from lambdaoid prophage Rac.
GN STFR OR B1372.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick J.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=9725157; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizoduchi K., Mori H., Mori T., Motomura K.,
RA Nakae S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map";
RL DNA Res. 3:363-377 (1996).
CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000234; AAC74454.1; ALT_INT.
CC EMBL: D90774; BA14966.1;
CC EMBL: D90775; BA14975.1;
CC EcoGene; EG13370; strf.
CC InterPro; IPR004089; Chmexis transd.
CC InterPro; IPR005003; Phage_fiber.
CC InterPro; IPR005068; Phage_fiber_2.
CC Pfam; PF03335; Phage_fiber_6.
CC Pfam; PF03406; Phage_fiber_2; 1.
CC Hypothetical protein; Fiber protein; Repeat; Complete proteome.

```
SQ SEQUENCE 1120 AA; 113779 MW; 542E5D71BE795B4 CRC64;
Query Match 5.7%; Score 138; DB 1; Length 1120;
Best Local Similarity 21.1%; Pred. No. 0.62;
Matches 106; Conservative 65; Mismatches 221; Indels 110; Gaps 17;

QY 3 INQSAQPPGVAMESFRFRTASDASLASSVRSVSTTSCRDQLAITDYLKHHVFAAHRFSVI 62
DB 110 VEEVARNASAVQNTAAAKKASDASDAREAAATHAA----- 146
QY 63 GSPDERDAALAHNEQIDALVETRANRLYSEG--ETPATIA-ETFAKAEKFDRLATTASSA 119
DB 147 ---DAADSARAASSTAGQAASQAASSAGSTASTKATEASAAABESSKSAATSAGA 203
QY 120 FENTPFAASVLQYMQPAINKGDMWLATPLKPLTPLISGALSGAMDQVTKWMDRARGDLH 179
DB 204 AKTSETNASASLQS-----AATSASTATTKASEAATSARDAAASK--EAAKSSET 251
QY 180 YLSTSPD-----KLHDMAYSVKHSHPALGRQV--VDMGIAVQTFSSALNVVRT 225
DB 252 NASSSASSAASATAAGNSAKAAKTSETNRSSETAAGCSASAAAGSKTAAASASAAST 311
QY 226 VLAPALASRPSVQGVADFCVSTAGGLVANAGFGDRMLSVQSGRDQLRGCAFVLGMKDKEPK 285
DB 312 SAGQASASATAGKAEASASASTATTKAGEATEQASAAARS-----ASAAKTSETN 364
QY 286 AALSBETDMLDAIKASYSYGAALNAGKRMAGLPLDVATDGLKAVSLVSATSLTKNG 345
DB 365 AKASETS--AESKTAASASSASSASSASS--ASASKDEATQASAAKS--SATTASTKA 418
QY 346 LALAGYAGVSKLQWAKYNTIDTSATKAAVSQLSNLVSQVGFAGWTTAGLATDPAVKA 405
DB 419 TEAAG-----SATAAQSKSTA-----ESAATRAETAAKRA 449
QY 406 ESFIQDKVKSTASSTTSYVADQTVKLAKTVKMDSGEATISSTGASLRSTVNNLRHSAP 479
DB 450 EDIASVALEDASTTKGI-----VOLSSATNSTS--ETLAATPKAVKSYADNAEKRLQDQ 504
QY 465 --ADIEEGG-----ISAFSRSE 479
DB 505 NGADIPDKGCFLLNNINAVSKTD 526

RESULT 2
CH60 LACHE STANDARD; PRT; 540 AA.
AC O68324;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROL OR MOPA OR GROEL.
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
ON NCBI_TaxID=1587;
RN [1]
RC SEQUENCE FROM N.A.
PC STRAIN=LH212;
RX MEDLINE=98439363; PubMed=9766226;
RA Broadbent J.R., Oberg C.J., Wei L.;
RT "Characterization of the Lactobacillus helveticus groEL operon.";
RL Res. Microbiol. 149:247-253(1998).
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF031929; AAC29004.1; --
DR HSP; P06139; IGRLL
DR InterPro; IPR001844; Chaprin Cpn60.
DR Pfam; PF00118; Cpn60_TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 540 AA; 57638 MW; 4257DD845FC7C4B4 CRC64;

Query Match 5.7%; Score 137; DB 1; Length 540;
Best Local Similarity 20.8%; Pred. No. 0.28;
Matches 109; Conservative 72; Mismatches 198; Indels 144; Gaps 24;

QY 21 ASDASLASSVRSVSTTSCRDQLAITDYLKHHVFAAHRFSVI---GSPDERDAALAHNE 76
DB 2 AKDIKFSENARRSL---LKGVDKLDATVKTITGPKGRNVVLEQSYGNPDITNDGVTTAK 57
QY 77 QID-----ALVETRANRLYSEGETPATIAETFAKAEKFDRLATTAS-----S 118
DB 58 SIELKDYENMGAKLVAEAAOKTNDIAGDGTATTATVLTQAIAREGMKNVGTAGANPVGIRR 117
QY 119 AFENTPFAAASVLQYMQPAINKGDMWLA--TPLKPLTPLISGALSGAMDQVQ----- 167
DB 118 GIEKATKAAVDELHKISHKVESKDQIANVAASVSSASKEIGALITADAMEKVGHGCVITIED 177
QY 168 -----TKWMDRARGDL--HYLSTSPDKLH-----DAMAVSVKHSHPALGR 205
DB 178 SRGINTELSVVEGQMPRGVLSQVMTDNDKMEADLNPVILITDKKISNIQDILPLL-Q 236
QY 206 QVDMGIAVQTFSSALNVVRLAPALAS--RPSVQGVADFCVSTAGGLVANAGSGD-RML 262
DB 237 EIVQGG-----KSLIIADDITGEALPTLVINKIRGT--FNVA---VKAPGFGDRRKA 285
QY 263 SVQSRDQLRGCAFV---LGMKDKEPKAALSEETDMLDAYKAIRKASYSYGAALNAGRMAG 319
DB 286 QLQDIAALTGGTVITEDLGLKDKTK----- 311
QY 320 LPLDVATDGLKAVSLVSATSLTKNGLALAGGVAG-----VSKLQWAKYNTIDTSAT 371
DB 312 -----IDQLGQARRI---TVTKDSTTIVGG-AGSKEAIDERVDTIRKQIEDSTSDFDK 360
QY 372 KAAVSQSLNLVSGVGFAGWTTAGLATDPAVKAESFIQDKVKST-ASSTTSYVA---DQ 427
DB 361 KQQLERLAKUTGGVAVI---HVGAAATETELKERRYRIEDALNSTRAAVDEGYVAGGTA 416
QY 428 TVKLAKTVKMDSGEAI--SSTGASLRSTVNNLRHSAPAEADIEE 469
DB 417 LVNVEKAVREVKGSETTDEQTGINI-----VLRLSAPVRQIAE 454

RESULT 3
FLJB SALAE STANDARD; PRT; 500 AA.
AC P52615;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Phase-2 flagellin.
GN FLJB.
OS Salmonella abortus-equi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
ON NCBI_TaxID=607;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=931139045; PubMed=84231149;
```


<hr/>					
DE	Hypothetical protein TC0868.				
GN	TC0868.				
OS	Chlamydia muridarum.				
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.				
OX	NCBI_TaxID=83560;				
RN	[1]				
<hr/>					
SEQUENCE FROM N.A.					
RC	STRAIN=MoPn / Nigg;				
RX	MEDLINE=20150255; PubMed=10684935;				
RA	Read T.D., Brinham R.C., Shen C., Gill S.R., Heidelberg J.F.,				
RA	White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,				
RA	Linhir K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,				
RA	Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,				
RA	Eisen J., Fraser C.M.;				
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia				
RT	pneumoniar39";				
Nucleic Acids Res.	28:1397-1406(2000).				
-!	SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0808/CT579/TC0868				
FAMILY.					
<hr/>					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
<hr/>					
EMBL; AE002353; AAF39664.1; --					
TIGR; TC0868; --					
Hypothetical protein; Complete proteome.					
SEQUENCE 436 AA; 43891 MW; 43E21C9924CD7A48 CRC64;					
<hr/>					
Query Match	5.4%; Score 129; DB 1; Length 436;				
Best Local Similarity	20.1%; Pred. No. 0.64;				
Matches 85; Conservative 67; Mismatches 170; Indels 100; Gaps 14;					
<hr/>					
QY	66 DERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLL-----ATTASSAF 120	:	:	:	:
DB	30 DXNGAAAKTQQQEASFEDLIQDSQGTGSKASKASTSQTSKGSEKAQKSSGYTTTTSVA 89	:	:	:	:
<hr/>					
QY	121 ENTPFAAASVLQWQPAINKGDWLATPLKPLTLPLISGAL----SGAMDQVGTKMMDRARG 176	:	:	:	:
DB	90 QASOTATAQAVRGARGSGFNNDGSASILPSPSPNPVNGVLKKMGTLALMGLIMTLILAQA 149	:	:	:	:
<hr/>					
QY	177 DLHLYSTSPDKLDAMAVSVKRHSPALGRQVV---DWGIADVTFESALNVTVTLAPALAS 233	:	:	:	:
DB	150 SAKSWSS-----SFQQOQNQIQOVAMAPEIGNAIRTOAHAAATELOAQKSL 198	:	:	:	:
<hr/>					
QY	234 RPVSQGVADVFVGSTAGLV--ANAGFGDRMLSVSQRDLGGFVLGMKDKEPKAALSEET 292	:	:	:	:
DB	199 ISGITNIVGFVSVGGIILSAKSILG-----GLK-----SAAFNET 235	:	:	:	:
<hr/>					
QY	293 DWLDAYKAIKSAYSYGAAINAGKRMAGLPDLDVTDGLKAVRSLSVTSATSLTKNGLALAGGY 352	:	:	:	:
DB	236 -----ATAASSATSAAKTAVNALDDVAN-----VAATAGTK-----AASGAA 273	:	:	:	:
<hr/>					
QY	353 AGVSKLKQWKATKNITDSATKAAVSQISNLVGSVGVFAGWTAGTAGLTDPDAVKKAESFTQDK 412	:	:	:	:
DB	274 SAASSAATKLTONMTDTASKTLSQTASKATG--GLF----- 307	:	:	:	:
<hr/>					
QY	413 VKSTASSTSYVDVQTVKLAQTKVDMWSGEAISSTGASLRS--TVNNLRHSAPEAIDEEG 470	:	:	:	:
DB	308 --GNAINTPNW--SEKISRGLNVVKTQGGRAAQAGRALSSAMSISQMVGHLTGAGVDITG 364	:	:	:	:
<hr/>					
QY	471 GI 472				
DB	365 GI 366				
<hr/>					
RESULT 6					
ID	SPCP RAT	STANDARD;	PRT;	2388 AA.	

AC O9QW8; 088197; Q9S68;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Spectrin beta chain, brain 2 (Spectrin, non-erythroid beta chain 2)
 DE (Beta-III spectrin) (SPNB-3) (Beta Spili sigma 1) (Spectrin-like
 DE protein GTRAP41).
 GN SPTBN2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=98340016; PubMed=9675416;
 RA Ohara O., Ohara R., Yamakawa H., Nakajima D., Nakayama M.;
 RT "Characterization of a new beta-spectrin gene which is predominantly
 RT expressed in brain.";
 RL Brain Res. Mol. Brain Res. 57:181-192(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=98359639; PubMed=9704016;
 RA Sakaguchi G., Orita S., Naito A., Maeda M., Igarashi H., Sasaki T.,
 RA Takai Y.;
 RT "A novel brain-specific isoform of beta spectrin: isolation and its
 RT interaction with Munc13.";
 RL Biochem. Biophys. Res. Commun. 248:846-851(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Rothstein J.D., Jackson M.;
 RT "GTRAP41 - a cytoskeletal modulator of EAAT4 glutamate transporter.";
 RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: Probably plays an important role in neuronal membrane
 CC skeleton.
 CC - TISSUE SPECIFICITY: Abundantly transcribed in the brain. Neurons
 CC are the predominant cell-type to express the gene. Found
 CC abundantly in Purkinje cells.
 CC - SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
 CC - SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
 CC - SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC - SIMILARITY: CONTAINS 17 SPECTRIN REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB008551; BAA32699.1; -;
 DR EMBL; AB001347; BAA32473.1; -;
 DR EMBL; AF225960; AAG28596.1; -;
 DR HSSP; Q01082; IBKR.
 DR InterPro; IPR001589; Actbind actnin.
 DR InterPro; IPR001715; Calponin-like.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00307; CH; 2.
 DR Pfam; PF00435; spectrin; 17.
 DR PRINTS; PR00683; spectrin; 17.
 DR SMART; SM00033; CH; 2.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00150; SPEC; 17.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00020; ACTININ_2; 1.
 DR PROSITE; PS00021; CH; 2.
 DR PROSITE; PS00003; PH_DOMAIN; 1.
 KW Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein.

FT DOMAIN 1 278 ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 57 161 CH 1.
 FT DOMAIN 176 278 CH 2.
 FT REPEAT 305 415 SPECTRIN 1.
 FT REPEAT 425 529 SPECTRIN 2.
 FT REPEAT 531 639 SPECTRIN 3.
 FT REPEAT 641 745 SPECTRIN 4.
 FT REPEAT 747 850 SPECTRIN 5.
 FT REPEAT 852 956 SPECTRIN 6.
 FT REPEAT 958 1063 SPECTRIN 7.
 FT REPEAT 1065 1170 SPECTRIN 8.
 FT REPEAT 1172 1276 SPECTRIN 9.
 FT REPEAT 1278 1381 SPECTRIN 10.
 FT REPEAT 1383 1486 SPECTRIN 11.
 FT REPEAT 1488 1586 SPECTRIN 12.
 FT REPEAT 1588 1692 SPECTRIN 13.
 FT REPEAT 1694 1799 SPECTRIN 14.
 FT REPEAT 1801 1905 SPECTRIN 15.
 FT REPEAT 1907 2011 SPECTRIN 16.
 FT REPEAT 2013 2075 SPECTRIN 17.
 FT DOMAIN 2218 2328 PH.
 FT CONFLICT 326 338 VTL -> GTF (IN REF. 3).
 FT CONFLICT 543 543 L -> F (IN REF. 2).
 FT CONFLICT 608 608 D -> G (IN REF. 3).
 FT CONFLICT 887 887 L -> P (IN REF. 3).
 FT CONFLICT 908 908 V -> I (IN REF. 3).
 FT CONFLICT 948 948 D -> G (IN REF. 3).
 FT CONFLICT 1156 1157 EL -> GA (IN REF. 2).
 FT CONFLICT 1194 1194 F -> V (IN REF. 2 AND 3).
 FT CONFLICT 1555 1555 G -> R (IN REF. 3).
 FT CONFLICT 1769 1769 R -> W (IN REF. 3).
 SQ SEQUENCE 2388 AA; 271062 MW; 3EC896AF0665F19 CRC64;

Query Match 5.3%; Score 128.5; DB 1; Length 2388;
 Best Local Similarity 21.8%; Pred. No. 6.1; Mismatches 168; Indels 143; Gaps 18;

Matches 101; Conservative 51; Mismatches 168; Indels 143; Gaps 18;
 QY 35 STTSCRDLOAITDYLKHHYFAHRFVIGSPDERDALAHNEQIDA-----LVETRAN 87
 DB 666 SATGTGDLTCVLRLLKHH--TALRGEMSRGLPKLTLEGQQLVABGHGCAQAOSTRA 723
 QY 88 RLYSEGETPATTAETPAKAEKFDRLATTASSAFENTPPAAASYLVQYQAPINKDMLATP 147
 DB 724 ELQAWERLEALAE-----ERAQRLA-----QAASLYQPADANDMEAWLVDA 766
 QY 148 LKRLTLFISALSGANDQYGTAKMDARGLDHLTSPDKLHMAAVSVKGRHPALGRQV 207
 DB 767 LR---LVSSPEVG-HDEFTQALARQ-----HRLAEERIRARRP----- 801
 QY 208 VDMGIAVQFSAINVRTVLAAPALASRPVQAVDPGVSTAGLVANAGPGR----- 260
 DB 802 -----TLDALREDAALPPALSHTEVQGRVPLTEGHYTEELQAPAGERARALEALAA 853
 QY 261 ---MLSVQSRDQLRGAFVYGMKDEPKKALSEETDMLDAYKAIKSASYSGAALNAGKRM 317
 DB 854 FYTMLS-----EAGACGLWVEKE-----QMLN----- 876
 QY 318 AGPLVDVADGLKAVASLYSATSITKXGLACGAVGSKLQMAATKNTTDSATKAAVSQ 377
 DB 877 -GLALPERLEDEVQORETLEPENMALA-----ARVTAVSDIAEQLLKASPPG 925
 QY 378 LSNLVGS-----VQVPAWTTAGLATDPAYK-----KAEFTQDKYKSTASS- 419
 DB 926 KDRITGTQEQLNORWQOQFSLADGKPAALTSALSIQNHLECTETAMMEKTKVIESIQ 985
 QY 420 -----TTSYVADQTVLAATVQKMSGEAISSTGASLRSTVNNL 457
 DB 986 DLGNDLAVGLAQR-KLAGTERNL--EALISARVGBELTQSANML 1025
 RESULT 7
 NCBJ_ALICXX
 ID_NCBJ_ALICXX STANDARD; PRT; 397 AA.

CH60_BATCHD STANDARD; PRT; 544 AA.

ID CH60_BATCHD STANDARD; PRT; 544 AA.

OS0305; Q9KFC3;

30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).

GN GLOL OR MOPA OR GROEL OR BH0562.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.

NCBI_TaxID=86665;

[1]

SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RC MEDLINE=97141316; PubMed=8987660;

RA Xu Y., Kobayashi T., Kudo T.;

RA "Molecular cloning and nucleotide sequence of the groEL gene from the

RT alkaliphilic Bacillus sp. strain C-125 and reactivation of thermally

RT inactivated alpha-glucosidase by recombinant GroEL.";

RL Biosci. Biotechnol. Biochem. 60:1633-1636 (1996).

[2]

SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RC MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;

RA "Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis.";

RL Nucleic Acids Res. 28:4317-4331 (2000).

CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND

CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS

CC CONDITIONS (BY SIMILARITY).

CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF

CC 7 SUBUNITS (BY SIMILARITY).

CC -1- INDUCTION: BY HEAT SHOCK.

CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

or send an email to license@isb-sib.ch).

EMBL; D55630; BAA09494.1; --

DR EMBL; AF001508; BAB04281.1; --

DR HSP; P06139; 1GRL.

DR InterPro; IPR001844; Chaprin Cpn60.

DR InterPro; IPR002423; Cpn60/TCP-1.

DR Pfam; PF00118; cpn60_TCP1.1.

DR PRINTS; PR00298; CHAPERONIN60.

DR PRINTS; PR00304; TCOMPLEXTCPL.

DR PROSITE; PS00296; CHAPERONIN CPN60; 1.

DR Chaperone; ATP-binding; Heat shock; Complete proteome.

FT CONFLICT 407 407 E -> VK (IN REF. 1).

FT CONFLICT 421 421 MISSING (IN REF. 1).

FT CONFLICT 500 500 A -> VP (IN REF. 1).

FT CONFLICT 512 512 MISSING (IN REF. 1).

SQ SEQUENCE 544 AA; 57403 MW; C769ED81C903C971 CRC64;

Query Match 5.1%; Score 123; DB 1; Length 544;

Best Local Similarity 18.9%; Pred.No.2;

Matches 104; Conservative 75; Mismatches 192; Indels 180; Gaps 23;

QY 21 ASDASLASSSVRSVSTTSRCDLQAITDYLK-----HHVFAHRFSVIGSP----- 65

DB 2 AKDIFSEDARRSM---LRGVKLDADVKTGLPGKRVNVLKFF---GSLPTNDGVT 54

QY 66 -----DERDALAHNEQIDALVETRNRLYSGEETPATIAETFAKAKFDRLATASAF 120

DB 55 IAKIEIELEDFENNGAKLVAEASKTNDIADGDTTVAIV---LAQAMIREGLKNVTGSA- 110

QY 121 ENTFFAAAVLYQWOPAIKNGDWLAPLKEPLTPLIS-GALSGAMDQV---TKWMDRARG 176
 DB 111 -NPMVIRKGIKAKQVAVBELSKISKIEGKDSIAQVAALISSADDEVKXIIAEMERVEN 169
 QY 177 D-----LHYISTSPDKLH-----DAMAVSVK 197
 DB 170 DGVITTEESKGFSTELLEVNGMOPDRGYAPYVVTSDSKMEAVLDNPFYLIITPKKLSNIQ 229
 QY 198 RHSPALGRQVVDGIAVQVTSALNVVTVLAPALASPSVQAVDFGVSTAGGLVAN--- 254
 DB 230 EVLPVL-EQVVOGSKPL-----LIADVGEALAT-----LVWVKLR 266
 QY 255 -----AGFGDRMLSVQSRDQLRGAFVLMKDKPKKALSEPTDWLDAYKAKSA 304
 DB 267 GFENAVAVKAPFGDR-----RKMLBDI----- 290
 QY 305 SYSGAALNAGKRAA-GLPLDVAIDGLKAVASLYSATSITKXGLALAG-----YAGVS 356
 DB 291 ----ALITGSEVITTEDGLDLKSNANITQL-GRASKVVTENTTIVGAGESDKIARVN 345
 QY 357 KLOKMAKNTDTSATKAASQLSNLVGSVGVFAGWTTAGLATDPAVKKASFIQDKVKT 416
 DB 346 QIRAQIEETTSDEPKELQERLAKLAGVAVL---KVGATETEMKERTLRLEDALNST 401
 QY 417 ASSTTSYV---ADQVYKLAKTVDMSGEAISSTGA-----SIRSTVNNLRHRSAPFADI 467
 DB 402 RAABEIVAGGGTALVNVKIVKVSIGAEGBATGVNIVRALIEPPIQAHNAAGLEGSV 461
 QY 468 -----EEGG 471
 DB 462 IVERLKKEAG 472

RESULT 9
 HPR_MOUSE STANDARD; PRT; 1068 AA.
 AC 9JUKY5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Huntingtin interacting protein 1 related (Hdpl-related).
 GN HDPLK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RC MEDLINE=20082868; PubMed=10613908;
 RA Engvister-Goldstein A.E.Y., Keseele M.M., Chopra V.S., Hayden M.R.,
 RA Drubin D.G.;
 RT "An actin-binding protein of the Slaz/Huntingtin interacting protein 1
 RT family is a novel component of clathrin-coated pits and vesicles.";
 RT J. Cell Biol. 147:1503-1518(1999).
 CC -1- FUNCTION: COMPONENT OF CLATHRIN-COATED PITS AND VESICLES, THAT MAY
 CC LINK THE ENDOCYTIC MACHINERY TO THE ACTIN CYTOSKELETON.
 CC -1- SUBUNIT: INTERACTS WITH ACTIN AND HUNTINGTIN INTERACTING PROTEIN 1
 CC (HDPL). DOES NOT INTERACT WITH HUNTINGTIN (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, MEMBRANE-ASSOCIATED PROTEIN,
 CC MAINLY LOCALIZED AT THE ENDOCYTIC COMPARTMENTS AND IN THE
 CC PERINUCLEAR REGION.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED, EXPRESSED AT LOWER LEVELS IN
 CC SKELETAL MUSCLE AND HEART. THE LEVEL OF EXPRESSION DOES NOT CHANGE
 CC APPRECIABLY DURING DEVELOPMENT.
 CC -1- DOMAIN: THE TALIN-LIKE DOMAIN BINDS F-ACTIN.
 CC -1- SIMILARITY: BELONGS TO THE SLAZ FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF221713; AAF34662.1; -.
 DR MGD; MG1:1552504; HDPL.
 DR InterPro; IPR001026; ENTH.
 DR InterPro; IPR002558; ILWEO.
 DR Pfam; PF01417; ENTH; 1.
 DR Pfam; PF01608; I_LWEO; 1.
 DR Prodom; PD011820; ILWEO; 1.
 DR SMART; SM00273; ENTH; 1.
 DR SMART; SM00307; ILWEO; 1.
 DR Actin-binding; Coiled coil.
 FT DOMAIN 77 80 POLY-SER.
 FT DOMAIN 311 316 POLY-GLU.
 FT DOMAIN 346 644 COILED COIL (POTENTIAL).
 FT DOMAIN 800 1068 TALIN-LIKE.
 SQ SEQUENCE 1068 AA; 119484 MW; 97CE9D92CDF5DB1 CRC64;
 Query Match 5.0%; Score 121.5; DB 1; Length 1068;
 Best Local Similarity 22.7%; Pred No. 5.7; Indels 153; Gaps 29;
 Matches 119; Conservative 74; Mismatches 178; Indels 153; Gaps 29;
 QY 1 MINOQAQPPGVAMBSFRTASDASLASSVRSVSTSCRDLOAITDYLNHVPAAHRS 60
 DB 505 MKMEQSDQLEKRLKRLAAGELAAQAL-SRTQSGSELSSRLDTLN-----AEKEA 558
 QY 61 YIGSPDERDAALAH-----NEQIDALVETPRANKLYSEGETPATIATTPAKAEKF----- 109
 DB 559 LSGVNRQREAEELAAQSLVREKEBALSOQRSSQKGLRGQLAKEKEQOGLRQKLLD 618
 QY 110 DRLATTAASFENTPPAAASVLYQWOPAIKNGDWLAPLKEPL-----TP--LISGALSG 161
 DB 619 EQLAVLRS-----AABAENILQDAVSKLD-----DPLHRLCTSPDVLVSRN-QA 663
 QY 162 AMDQVTKWMDRARGDLHYLSTSPDKLHDMAVSVKHSRHPALGRQVVDGIAVQVTS--- 218
 DB 664 ALDSVS-----GLEQHTQYGLASEDA--SALVALALTRFHLADTVN-GAATSHLAPMD 716
 QY 219 -----ALNVV-----RTVLAAPALAS--RPSVQAVDPGVSTAGGLVAMNG 266
 DB 717 PADRLMDTCRECGARALBELVQLODQTVLRRAQPSIMRAPLQGIILQIGDDL----- 767
 QY 257 FGDRMLSVQSRDQLRGARAVLGMKDKPKKALSEPTDWLDAYKAKSA-----SYSGAA 310
 DB 768 ---KPSLDVROBELG-----AVNDKE-MAATSAITE--DAVRIDMMSQARHSSGVK 816
 QY 311 LNAKRMAGLPIDVATDGLKAVASLY-SATSLTKNGLALAGYAGVSKLOKATKNI--T 367
 DB 817 LEVNERI-----LNSCDLMKATRLVMTSTLQKE---IVESGRGAATQDEFPYAKSRMT 869
 QY 368 D---SATKAASQLSNLVGSVGVFAGWTTAGLATDPAVKKASFIQDKVY-----S 415
 DB 870 EGLISASKA-----VGMGATQL-----VESADPVVLHMKYEBELVCSHE 909
 QY 416 TSSTTSYVADQTVK-----LAKTYKMSGEAISST 446
 DB 910 IAASTQALVAASKVKANKNSPHLSRLQECSTRVNERANVAVST 953

RESULT 10
 YM96_YEAST STANDARD; PRT; 1140 AA.
 ID YM96_YEAST
 AC 004893;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 113.1 kDa protein in PR55-PEF4 intergenic region.
 GN YMR317W OR YM9924.09
 OS Saccharomyces cerevisiae (Baker's yeast).
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

```

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z54141; CAA90835.1; -
DR SGD; S0004936; YMR317W.
KW Hypothetical protein; Repeat.
SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;
Query Match 5.0%; Score 120.5; DB 1; Length 1140;
Best Local Similarity 19.8%; Pred. No. 7.2;
Matches 118; Conservative 84; Mismatches 233; Indels 161; Gaps 26;
Qy 12 GVAMESFTASDAS--LASSVRSV-----STTSCRDQLAIDYDKHVFPAHRFSVIGSP 65
Db 203 GSAVTSVGTSDASKEVFSSTSDVSSLLSTSSPASSTISETLP---FSSTILSITSP 259
Qy 66 DERDAALAHNEQIDALVETANRLYS-----EGETPATIA-----ETFAKA 106
Db 260 VSSEAPATSSVSSEASSTSSVSSEAPLATSSVSSEAPSTSSVSSEAPSTSSS 319
Qy 107 EKFDRLATAGSAFENTPFAAASLYQMPAINKGDWLA-----TLPKPLTP 153
Db 320 VSSEISSTSSVSSEAPLATSSVSSEAPSTSSVSSEISSTSSVSSEAPLATSSV 379
Qy 154 LISGALGAMQVGYKMDRANGDLHYLSTSPDKLHDAMAVS-KRHPALGRQV----- 207
Db 380 VSSEAPSTSSVSSEAPSTSSVS--SEAPSTSSVSSEISSTKSSVMSSEVSATS 437
Qy 208 -----VDMGTAVQTFGALN--VVRTVLA-----PALASRPSVOGAV 241
Db 438 SLVSSEAPSAISSLASSRLFSKNTSVTLVATEASSVTSLSRPSSTLASNSISSL 497
Qy 242 DFG-----VSTAGGLVANAGFGDRMLSVQSR-----DOLRGAFVLGMKD---KEP 284
Db 498 STGYNSTVSTTTS--AASSTLGSKVSSNSRWATSKTSTSSDLKSSVIFGNSSTVTTSP 556
Qy 285 KAALSEETD-----WLD-----AYKAIKASYSYGALNAGKRMAGLPLDVATDGLKAVRS 334
Db 557 SASISLTASPLPSVMSDITSSPASISSNASSAPSDNNSTIASASLIVTKNSVVSS 616
Qy 335 LV-----SATSLTKGLAL-----AGGAGVSKLQKMAT 363
Db 617 IVSSTTSSETNESNLATSSSTSLSNKATARSLSNTSNTASNVPTGTFSSMSTSVIT 676
Qy 364 KNITDSATKAASQSLNVLGVGVFAGW-----TTAGLATDPVAKKAES-----F 408
Db 677 PGFSTSSASLAIN--STVVSS--SLAGYSFSTPESPSTTSLTSEAPSTVSSMTTSP 732
Qy 409 IODKVKSTAS-STTGYVADQTVKLAKTKVMSGEAISSTGASLRSTVNNLHRSAP 463
Db 733 INNSTARPSPSTASFITESTSSIS-SVPLASGDVTSLLA-----HNLTTFSAP 781
RESULT 11
MSB2 YEAST
ID MSB2_YEAST STANDARD; PRT; 1306 AA.
AC P32334;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)

```

```

DE MSB2 protein (Multicopy suppression of a budding defect 2).
GN MSB2 OR YGR014W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92383951; PubMed=1514328;
RA Bender A., Pringle J.R.;
RT "A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence
RT defect.";
RL Yeast 8:315-323(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT chromosome VII.";
RL Yeast 13:1077-1090(1997).
CC -!- PTM: O-GLYCOSYLATED IN THE SER/THR-RICH REGIONS (PROBABLE).
CC -!- SIMILARITY: SOME, TO YEAST HKR1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M77354; AAA34798.1; -
DR EMBL; Z72799; CAA96997.1; -
DR PIR; S25370; S25370.
DR SGD; S0003246; MSB2.
KW Transmembrane; Glycoprotein; Repeat.
FT DOMAIN 698 816
FT REPEAT 698 714 1.
FT REPEAT 715 731 2.
FT REPEAT 732 748 3.
FT REPEAT 749 765 4.
FT REPEAT 766 782 5.
FT REPEAT 783 799 6.
FT REPEAT 800 816 7.
SQ SEQUENCE 1306 AA; 133114 MW; 67D5D984D5CA4A6D CRC64;
Query Match 5.0%; Score 120; DB 1; Length 1306;
Best Local Similarity 20.2%; Pred. No. 9.1;
Matches 117; Conservative 79; Mismatches 235; Indels 148; Gaps 24;
Qy 14 AMESFRITASDASLAS-----SSVRSVSTTSCRDQLQAIDYDKHVFPAHRFSVIGSPDER 69
Db 142 AASSVNAASDVSTASVPIDTSANSIPFTTTSNIETTT-----SAPLTS 185
Qy 70 AALAHNEQIDALVETANRLYSGETPATIAETPAKAEKFDRLATTASSAFENT-PPAAA 128
Db 186 TPLISTSTMSA-----ADNVFSSANPISASLTITDSSSEFDQTSTAGAIPOVQSSADFS 240
Qy 129 SVLQYMPAINKGDWLAETPLKPLTSLGALSGAMQVGT-KTMDRARGDLHYLSTSPDK 187
Db 241 SEILVQSSAD-----FSSPSSTTTDIS--LSAAPTQSSSSSTTSAALPVGSTDVDG 293
Qy 188 LHDAMAVSVKRHSHPALGRQVVDMGIAVQTFESALNV-----VRT--- 225
Db 294 SSASPVVSMASGAAGQIASSSSSTDNPTMSETFSLTSTEVDGSDVSVTSALLSAPLQTS 353
Qy 226 -----VLAPALASRRPSVQAVDFGVSTAGGLVANAGFGD-----RMLS 263
Db 354 NSFISVPSVPSFVPS-QSSSDVASSSTANVVSS--FSDIPPQTSTSGSVVSAQSALA 411
Qy 264 VQRDLQRGGAFVLGMKDKPKAALSSETDWDLAKAISKASYSYGAALNAG-KRMAGLPL 322

```

```

Db 412 FQSTEVYGAASSTMSLSLSTSLQSTT--LDSSSLASSASSSDLTQVGSSTASIPL 469
QY 323 DVATDGLKAVRSL-----VSATSLTKNGLALAGYAGVSKLOKMA 362
Db 470 LSASEQASTSSFFSVSPSVFVPSQSSDVASTSAPSVSSSFYSLSQAGSSMTNPS 529
QY 363 TKNI-----TDSATKAAVSQLS-NLVGSVGVFAGMTAGLTADPAVKKASFIQDKYKST 416
Db 530 SSTIVSSSTGSSSESAASTATLSGSSSTY---MAGNLQSQP--PSTSLSE---SQ 581
QY 417 ASSTTSVVAQTVLAK---TVKMSGEA---ISSTGA-----SLRSTVNNLRKRSAPEA 465
Db 582 ATSTSAVALASSSVSTSPYTTAGASTASLSTSTSAETSQVSYSQSTTALQTSFPASS 641
QY 466 DIEEGC-----ISAFSRSETPFL 484
Db 642 STTEGSETSSGCFSTSSVLVQMPSSISSEFSPQTTTQM 680

RESULT 12
HLVA_SERMA
ID HLVA_SERMA STANDARD; PRT; 1608 AA.
AC P15370;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.
OS SHLA.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Serratia.
OX NCBI_TaxID=615;
OX (1)
RN RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RX STRAIN=SN8;
RX MEDLINE=88257037; PubMed=3290200;
RA Poole K., Schisbel E., Braun V.;
RT "Molecular characterization of the hemolysin determinant of Serratia
marcescens.";
RL J. Bacteriol. 170:3177-3188 (1988).
CC -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
DEFINED.
CC -1- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
REQUIRES SHLB FUNCTION.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: TO P. MIRABILIS HEMOLYSIN (HPMA).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M22618; AAA50323.1;
DR PIR: A28182; A28182.
KM Hemolysins; Toxin; Outer membrane; Signal.
FT SIGNAL
FT CHAIN 31 1608 HEMOLYSIN.
SQ SEQUENCE 1608 AA; 165078 MW; D6698476FE7DAD51 CRC64;

Query Match 5.0%; Score 119.5; DB 1; Length 1608;
Best Local Similarity 20.3%; Pred. No. 13;
Matches 120; Conservative 68; Mismatches 242; Indels 155; Gaps 21;

QY 8 QQPGVAMSEFRTPASDASLASSSVRSVTSQCDLOAITDYLKHHVPAARFVSITGSPDE 67
Db 708 QYRAGLRIETRRSEKTRTRENASSSSGSVK-LKAEKV---TSGSKLVA---DK 758
QY 68 RDAALAHNEQIDALVETRANRLYSEGETPATIAET-----FAKAEKFDRLATTASSAFEN 122

```

```

Db 759 GDASVSGN-----KYSFLAADDKTASNTQTKIGGFYYTGIDKLGSGVEAGYEN 809
QY 123 TTPFAASVLYQMPAINKGWMLATPLKPLTSLGSLGSGMDQVGTMMRBARGDLHY-- 180
Db 810 NKTQPOSSKAITSQSDVKGMLTTNARDKLTQ-----QGQHSVGVAGYQNNAGVDHLAA 863
QY 181 ---LSTPDLHDMAVSVGRHSPALGRQY---VDMGIAVQTSALNVTVLAPALASR 234
Db 864 ADTASTTTTQTDGVNIGAVVDYSAYTRPVRVAGRAKLDATGCVINDIGIGAPNVGLD 923
QY 235 PSYQAVDFGVSTAGLIVAN---AGFEDRLSVQSRDQ----- 269
Db 924 IGAQGSSEKSSSSQAVSVQAGSIDINAKGEVDDQGTQYQASKAVNL/TDASHRSEA 983
QY 270 -----LGGAFVL-----CMKDEPRAALSEETDMDAYAI- 301
Db 984 AANRQEQSDRTGSAQVRYTTGSDLTVDAKGEGTQSNSSASQAVTGSIDPAANGIN 1043
QY 302 ---KSASYSGAALNAGKRM---AG--LPLDVATDGLKAVRSLVSATSLTKNGLA---- 347
Db 1044 VNNKQDAIYGTALNGRCKTAVNAGDITLDDASDKQSSRSRGFNVKASAKGFTADSK 1103
QY 348 -LAGYAGVSKLOKMATKNTTDSATRAAVSOLS-----NLVGSVG 386
Db 1104 NFGAGFGG-----GTHNGESSSTQVGNISGQCGVELKAGRDLTLQGTDVKSQGDVS 1156
QY 387 VFAGMTAGLATPAPVKAASFQDKYKSTASTTSVADQYTKLAKTVDMGSAISST 446
Db 1157 LSAQNKVALQMAESTQTRKESKLSGNIDLAGSSDS-----XKTKGNLSAG 1203
QY 447 GASLRSTVNNLRHRSAPV---ADIEEGC---ISAFSRSETPFL 485
Db 1204 GADDIATVN-----ESATERGATITASDGKTYLSANGKGDALHQ 1244

RESULT 13
HS7F_CABEL
ID HS7F_CABEL STANDARD; PRT; 657 AA.
AC P11111; P91135;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Heat shock 70 kDa protein F, mitochondrial precursor.
GN HSP-6 OR HSP70P OR C37H5.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderae; Caenorhabditis.
OX NCBI_TaxID=6239;
OX (1)
RN RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2.
RA Davidson S., Gilliam B.;
RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.
RN [2]
RN RP SEQUENCE OF 1-487 FROM N.A.
RX STRAIN=Bristol N2.
RX MEDLINE=89356251; PubMed=2766926;
RA Heschl M.F.P., Baillie D.L.;
RT "Characterization of the hsp70 multigene family of Caenorhabditis
elegans.";
RL DNA 8:233-243 (1989).
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

```
DR EMBL; U88315; ABB42371.1; -.
DR EMBL; X07678; CAA30525.1; -.
DR PIR; S03210; S03210.
DR PIR; B32475; B32475.
DR HSP; P04475; 1DG4.
DR WormPep; C37H5.8; CE08631.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; Hsp70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR ProDom; PD000089; Hsp70; 1.
DR PROSITE; PS00297; Hsp70; 1.
DR PROSITE; PS00329; Hsp70; 2.
DR PROSITE; PS01036; Hsp70; 3.
DR ATP-binding; Heat shock; Multigene family; Mitochondrion;
KW Transist peptide.
FT TRANSIT 1 27 MITOCHONDRION (POTENTIAL).
FT CHAIN 28 657 HEAT SHOCK 70 KDA PROTEIN F.
FT CONFLICT 138 138 V -> E (IN REF. 2).
FT CONFLICT 140 140 S -> P (IN REF. 2).
FT CONFLICT 167 167 KA -> NV (IN REF. 2).
FT CONFLICT 347 348 KA -> NV (IN REF. 2).
SQ SEQUENCE 657 AA; 70844 MW; 637A1636B6AB836 CRC64;

Query Match 4.9%; Score 119; DB 1; Length 657;
Best Local Similarity 21.5%; Pred. No. 4.3; Mismatches 162; Indels 168; Gaps 26;
Matches 107; Conservative 60;

QY 29 SSVRSVSTSCRDQAIYDYLKHVFAAHRSFVIGSPDPERDAALAHNEQIDALVETRANR 88
DB 9 SSARTIARSLMSARSLSDPKGHVIGID-----LGTNSCVSINEGKTPK 54

QY 89 LYSGE-----TPATIAETFAKAEF-----DRLATTASSAFENTPFAASVL--QYQPA 137
DB 55 VIENAEGRVTFSTFVAF--ADGERLVGAPAKRAQVATNSA---NTLFATKRLIGRRYEDPE 110

QY 138 INK-----CD-WLATPLKPLTPLISGALSGAMDQVGTMMDRARGDLHVL 182
DB 111 VKDKLVVYKIVKASNDGAWVEAGKVYSPSQVAF-----VLMKMKETAES---YLG 161

QY 183 TSPDKLHDAMAVSVKRHSPALGRQV--DMGIATVTFSAINVRTVLPALASRPSVOGAV 241
DB 162 TVNN-----AVTVPAYENDSORQATKAG---QISGLNVLVINEPTAA-----204

QY 242 DFGVSTAGLVANAGFGDRMLSVGRDQRLRGAFVLGKMKPKPAALSEETDWDIAKAI 301
DB 205 ----ALAYGLDKDA--GDKIIAVY---DLGGGTF-----DVSILEIQK--238

QY 302 KSASVSGAALNAGRMAGLPLDVAIDGLKAVRSLSVATSLTKNGLALAGGAGVYKQLQM 361
DB 239 -----GVFEVKS-----TNGDTFLGG-----254

QY 362 ATKNIITDSATKAASVQLSNLVSQVGFAGWTAGIATDPVAKAESFTQDKVKSTASTT 421
DB 255 -----EDFDHALVHL-----VGEFKKEQGVDTKDPQAMQRLREAEAKKCELSSTT 302

QY 422 S-----YVA-DQT-----VKLAKT-VKDSMGEAISSTGASLSSTVNNLHRAPENDI 467
DB 303 QTDINLPYITMDQSGPKHLNKLTRAKEQIVGDLIKETIETPCRKALHDAEYKVSQIADV 362

QY 468 -EEGGSIAFSRSETPFQ 483
DB 363 LLVGMRSRMPKVQATVQ 379

RESULT 14
METE MYCLE
ID METE MYCLE STANDARD; PRT; 760 AA.
AC 005564;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 5-methyltetrahydropteroylriglutamate--homocysteine-methyltransferase
DE (EC 2.1.1.14) (Methionine synthase, vitamin-B12 independent isozyme)
```

```
DE DE (Cobalamin-independent methionine synthase).
GN METE OR ML0961 OR MLCB33.14.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- FUNCTION: CATALYZES THE TRANSFER OF A METHYL GROUP FROM 5-
CC METHYLTETRAHYDROFOLATE TO HOMOCYSTEINE RESULTING IN METHIONINE
CC FORMATION (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 5-methyltetrahydropteroyltri-L-glutamate + L-
CC homocysteine = tetrahydropteroyltri-L-glutamate + L-methionine.
CC -!- COFACTOR: ZINC; BINDS ONE ION PER SUBUNIT (BY SIMILARITY).
CC -!- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.
CC -!- SIMILARITY: BELONGS TO THE VITAMIN-B12 INDEPENDENT METHIONINE
CC SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z94723; CAB08123.1; -.
CC EMBL; AL583920; CAC31342.1; -.
CC Leproma; ML0961; -.
CC InterPro; IPR002629; Methionine_synth.
CC Pfam; PF01717; Methionine_synth; 1.
KW Transferase; Methyltransferase; Methionine biosynthesis; Zinc; Repeat;
KW Complete proteome.
FT METAL 647 647 ZINC (BY SIMILARITY).
FT METAL 649 649 ZINC (BY SIMILARITY).
FT METAL 732 732 ZINC (BY SIMILARITY).
SQ SEQUENCE 760 AA; 82234 MW; 1E92FAS70B9ACC97 CRC64;

Query Match 4.9%; Score 118.5; DB 1; Length 760;
Best Local Similarity 21.1%; Pred. No. 5.6;
Matches 91; Conservative 66; Mismatches 173; Indels 101; Gaps 18;
```

```
QY 85 RANRLISEGTPATIAETFAKAEFDRLATASSAFENTPFAASVLQYMQPAINKGDWL 144
DB 28 RATEGYWAKETSRSELSVASTLRRDMWSDLAAAGLDVSP---VNTFSYYDQMLD-----79

QY 145 ATPKLPKPLTPLISGALSGAMDQVGTQM-----MDRARGDL-----HYLSTSPD 186
DB 80 -----TAFMLGALPARVAQVSDDLDOYFALARGNNDIKPLEMTKWFDTNYHYL-----127

QY 187 KLHDAMAVSVKRHSPALGRQVDMGIAVOTFS-----ALNVRTVLPALASRPSVOGAV 241
DB 128 -----VPEIEPA-----TTFSLNPGKILGELKEALQRIPIRPIVIGPV 166

QY 242 DF-----GVSTAGLVAN-----AGFGDRMLSVQSRD-----QLRGGAFLVGMKDKPKAA 287
DB 167 TFLLLSKINGGGGAPIQRLBELGIYCTLLSLAENGARWVQFDEPALVTDLSPDAPALA 226

QY 288 LSEBTDWLDAYK--AIKSASVSGAALNAGRMAGL---PLD-VATDGLKAVRSLSVATSL 341
DB 1
```

Db 227 EAYTALGSKSPKPAIVATYFC--NPGASLACIANTPIEALVDVPCGADISVAAPVE 283
 Qy 342 TKQGLALAGYAGVSKLOKRNATKNTDPAKVAVSQSLNVSQVGFAGWTAGLATDPA 401
 Db 284 LAGKTLVAGIVD-----RWMTDLESALSKATLGLSAAVAVTSCTLAVPY 334
 Qy 402 VKRAESTIQKVSSTASTSYVADQVYKATV---KDSGSAISSTGASLSTVNNLR 458
 Db 335 SLBETDLDNLRSMWLFAGAEKVA-EVYVLAFLRDGRDAVADEIAASNAVSRSDPR 393
 Qy 459 -HRAPEADIE 468
 Db 394 LHNQVRRRID 404
 RESULT 15
 LYTD_BACSU STANDARD; PRT; 880 AA.
 ID LYTD_BACSU STANDARD; PRT; 880 AA.
 AC P39848;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 GN Beta-N-acetylglucosaminidase precursor (EC 3.2.1.96).
 OS LYTD OR CWLG.
 OC Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OC NCBI_TaxID=1423;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RC MEDLINE=95020588; PubMed=7934877;
 RA Margot P., Mausel C., Karamata D.;
 RT "The gene of the N-acetylglucosaminidase, a Bacillus subtilis 168
 cell wall hydrolase not involved in vegetative cell autolysis.";
 RL Mol. Microbiol. 12:535-545 (1994).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AC327;
 RC MEDLINE=96036197; PubMed=7581999;
 RA Rashid M.H., Mori M., Sekiguchi J.;
 RT "Glucosaminidase of Bacillus subtilis: cloning, regulation, primary
 structure and biochemical characterization.";
 RL Microbiology 141:2391-2404 (1995).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RC MEDLINE=98044033; PubMed=9384377;
 RA Kunze F., Ogatawa N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bortolero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Bortolero R., Bortolero L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brueschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Comerton I.F., Cummings N.J., Daniel R.A.,
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmertson P.T.,
 RA Eutenei K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz S.Y., Fujita M., Fujita Y., Fuma S., Gallazzi A., Galleron N.,
 RA Gilm S.Y., Glaeser P., Goffeau A., Goldlighty E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogg S., Kumano M.,
 RA Kuriita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Neone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Portolillo G., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadele Y.,
 RA Sato T., Scanlon E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Solde B.,
 RA Sorokin A., Taconi B., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vamher P., Vassalotti A.,
 RA Viari A., Wambut K., Wedler E., Wedler H., Weltzenegger T.,

RA winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis.";
 RL Nature 390:249-256 (1997).
 CC -1- FUNCTION: CELL WALL HYDROLASE NOT INVOLVED IN CELL AUTOLYSIS. IT
 CC HYDROLYZES THE BETA-1,4 GLYCAN BOND BETWEEN THE N-
 CC ACETYLGLUCOSAMINYL AND THE N-ACETYLIMURAMOYL RESIDUES IN THE GLYCAN
 CC CHAIN.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
 CC unit in high-mannose glycopeptides and glycoproteins containing
 CC the -[Man(GlcNAc)2]Aen-structure. One N-acetyl-D-glucosamine
 CC residue remains attached to the protein; the rest of the
 CC oligosaccharide is released intact.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: secreted.
 CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC
 CC EMBL: U02562; AAA67857.1; -
 CC EMBL: D45048; BAA08089.1; -
 CC EMBL: Z29122; CAB15595.1; -
 CC Subtilisin, BG10455, lytd.
 CC DR InterPro: IPR002901, Amidase_4.
 CC DR InterPro: IPR003646, SH3_bac.
 CC Pfam: PF01832, Amidase_4f.1.
 CC SMART: SM00047; Amydase 4f.1.
 CC SMART: SM00287; SH3b; 1.
 CC Cell wall; Hydrolase; Signal; Repeat; Complete proteome.
 CC FM SIGNAL 1 27
 CC FM CHAIN 28 880
 CC FT DOMAIN 72 75 BETA-N-ACETYLGLUCOSAMINIDASE.
 CC FT DOMAIN 337 340 POLY-THR.
 CC FT DOMAIN 568 571 POLY-LYS.
 CC FT DOMAIN 439 473 POLY-ALA.
 CC FT REPEAT 1.
 CC FT REPEAT 479 513 2.
 CC SQ SEQUENCE 880 AA; 95553 MW; 2A912A478FCFCD1 CRC64;
 Query Match 4.9%; Score 118.5; DB 1; Length 880;
 Best Local Similarity 18.3%; Pred. No. 6.8; Indels 141; Gaps 15;
 Matches 90; Conservative 77; Mismatches 185;
 Qy 26 LAASSVRSVSTSCRDLOAITDYKHHVFPAHRFSYVSGSPDEBDALAHNEQIDA----- 80
 Db 11 LSAASLAFPMSSQAQAAATDLSLYKVERSNFTFSQASQAVLKEKDTGWDASVQAS 70
 Qy 81 ----LVETRAHRLYSEGETPATTAETPAKAKEPRLATTASSAFENTPFAASVLOYMOP 136
 Db 71 GTTVOAISASGHSEBAKAIDS-----GLAQGISRTGSSPVGSK----- 112
 Qy 137 AINKGMLAPLKLPLPLISGALSAMDVGTGMBRARDLHYLSTSPKLDMAVSV 136
 Db 113 -----OPYTSSGALSIG-----EKOANTLIKL 136
 Qy 197 KRSPALGRQVDMGIAVQFSLANVRVTLAPALASRPVQAVDFGVSTAGLVANAG 256
 Db 137 KQETVAG-----AVKAYGAAGYVNMVMTSDIDDETVAKLIIQ-SLAKQGISSSYQ 187
 Qy 257 FGRMLSVQSRDLRGAFVYLGKDKRPAALSEETDWT-----DAYKAIKASYS 307
 Db 188 PITHVSVTL-----IQGTIVGDSRAQIKAFKESGLQSLKEVYKGQYVTFYTAIS 244
 Qy 308 GAA-----LNAGRMGL-----PLDVATDGLKAVRSLVSTSLTKGCLAL 348
 Db 245 GEANAKTLQLQKOSTGTGTSYKSNOKTTFVESYVQSAVFKGLSTVKAIISQIKNTGV 304

Qy	349		AGVYGVSK-----LQMAT-----KNITDGAATKAASQLSNLVGSUV	387
	:	:	: :: :	:
D6	305		SGSYQQVGKTSVTNMGKITQOLKDIDFFKKKHWTSSVKTKTTTSAAYOI----	359
	:	:	:: :	:
Qy	388		PAGWTTAGLATDPAAVKAESFI-ODKV-----STASTTSYADOTVKLAIVDKMS	439
	:	:	:: :	:
D6	360		-TTAKILGEOANKAAAFAFKKYKAAKTAAGSTAENYOVLISEETSDOAKVTKGIN	415
	:	:	:: :	:
Qy	440		GEAISSTCASLR5 452	
	:	:	:: :	:
D6	416		ILKNQLSAKS 428	

Search completed: January 2, 2003, 15:19:24
Job time : 14.5098 secs

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent
A:Reference number: A97359; PMID:11743194

RESULT 4
G98240
hypothetical protein AGR_L1758 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C/Species: Agrobacterium_tumefaciens
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C/Accession: G98240
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
R/Reference number: A57359; PMID:11743194

A/Accession: G98240
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1009 <KUR>
 A/Cross-references: GB:AE007870; PIDN:AAK89449.1; PID:g15159312; GSPDB:GN00170
 C/Genetic: AGR L 1758
 A/Map position: linear chromosome

Query Match 6.2%; Score 149; DB 2; Length 1009;

Best Local Similarity 22.4%; Pred. No. 0.13; Matches 127; Conservative 77; Mismatches 228; Indels 136; Gaps 25;

```

QY 7 ACPGCVASFTASASIASSSVRSVSTTSCRDLOAI--TDYKH---HVFAHRSFVI 62
DB 118 ASQIDAMSTQIRKANSQVAGSSAQVATLSSDELALFTTDEKISANAAIAGLSAAI 177
QY 63 GSPDERDALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEK-----DRLATTA 116
DB 178 AGLDNAALTKSQIAAMSTQFNALTSGLATFSADDEVKASINKILAGLDVTKSTGN 237
QY 117 SSAPENTPPAAASVLTQMOPAINKGDMLATPKLPLPLISGALSAMDQGTMMDAAG 176
DB 238 IALSKAQVASTTQFPAAASTDQ-----IKALTSDOVAGLSA--QVATLSSD---- 284
QY 177 DLHYLSTSPDKLHDAMAVSVKHSPPALGRQVVDGIA-----VQ 215
DB 285 ELALFSTDEIKAGANAVA--GLSAAALALTTDAAALTKQIAGLSTQALNTSANIA 343
QY 216 TFSALNVRTVLAPALASRPSVQAVDFGVSTAGLVA-----NAGFDRMLSVQSRDQ 270
DB 344 TFSR-DEIKAIKSTKALA-----GLDVTKLTSGNIALTKQQAASLSTQFAMSTDQ 395
QY 271 RG--GAIVLGMKKEPKAALSEE-----TWMUDAYKAIKASYSYGALNA----- 313
DB 396 KALTSQVAGISSAQVATLSSDELALFTTDEISAFSANAAGLSTALALTTGNATALT 455
QY 314 GKRWAGL---PLDVATGKLVRSVLSATSLTKNGLALAGVAVSKLOKMATKNTTDSA 370
DB 456 KTQAGLSTQALNLTSGSLNTFSADVKAIKSTK--ALLAG-----LDVTKLSTGNVA--AL 507
QY 371 TKAAVSQLSNLVGSVGFAGWTT-----AGLATPBAVKKAS-----FTQDRKVS 415
DB 508 SKAQVSALSTT-----QFAMSTQIKALTSQVAGLSAQVATLSSDELALFTTDEIKA 562
QY 416 -----TASSTTSYVADQTVKLAKT--VKDMSGEAISSTGASLSTVNLRHSAP- 464
DB 563 IGANAVAGLSMAALALTTDVASALTKQIAGMSTQINAL-----TSANLATFSADAI 616
QY 465 -----ADIEGGISAFSRSE 479
DB 617 KAITTKALGGLDVTKLSTGNIALTKAQ 644

```

RESULT 5
 T3110
 extracellular matrix binding protein - Abiotrophia defectiva (fragment)

C/Species: Abiotrophia defectiva
 C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
 C/Accession: T3110
 R/Manganello, R.; van de Rijn, I.
 Infect. Immun. 67, 50-56, 1999
 A/Title: Cloning and characterization of emb, a gene encoding the major adhesin of Strep
 A/Reference number: Z20988; MUID:59081722; PMID:9864195
 A/Accession: T3110
 A/Status: preliminary; translated from GB/EMBL/DBD
 A/Molecule type: DNA
 A/Residues: 1-205 <MAN>
 A/Cross-references: EMBL:AF067776; NID:g3249002; PID:g3249003; PIDN:AD03320.1
 C/Genetic: emb

Query Match 6.0%; Score 143.5; DB 2; Length 205;

Best Local Similarity 22.2%; Pred. No. 0.79; Matches 123; Conservative 78; Mismatches 249; Indels 105; Gaps 23;

```

QY 3 INSOAQP-----PGVAMESFRTASDASLA---SSSVRSVSTTSCRD--LQAI-----TDY 48
DB 726 INEISQRPDLTREKQAFMDQVETARDAAKAVASANNQAVTSADQGLNAAVNNLPTPA 785
QY 49 LKHVFAHRSFVIGSPDERDALAHN-----EQIDAL-----VETRANRLYSEGETPA 97
DB 786 AKYPEALGH---VRQADAKQQAIRDNANLTLAEQADALRQVDAQTAALAINQHTNA 842
QY 98 TIAETFAKAEK-----DRLATTSAPENTPPAAASVLTQMOPAINKGDML-----A 145
DB 843 TLAKADSDGVKALINDINPQFRSKPAANOALEQV--AAAK-----RQAINNNQTLTBEKA 895
QY 146 TLAKPLPLISGA---LSGMDQVGTMMDBRAGDHLHYLSTSPDKLHDAMAVSVKHSPPA 202
DB 896 QALIQVDQALANKTQVQAANDNNGVNOAKTA--GTTAINNINPQGTOKQAQALIAIEAEQ 954
QY 203 LGRQVVDGIAVQTFPALNVRTVLAPALASRPSVQ-----GAVDFGVSTAGLVA 253
DB 955 AKLELQGRNDLTTERNNALADLTAKQAQAKAVQAARNTGVAKXGVAQIGINP 1014
QY 254 NA-----GFGDRMLSVQSRDQLRGAFVLGMKKEPKALISEETDWL--DAYKA 300
DB 1015 TAVVPEPARVAIDQAAARDEKAEFOANTKL-----TDEKRAAAIKVQDAARDAKAA 1065
QY 301 IKASASISG---AALNKGKRAAGPLDVATDGLKAV--RSVLSATSLTKNGL-----A 347
DB 1066 IDRAGSNGDVNNVNOGK-----AALQAIKALDSDQSPADTAQALQNAADAKKAA 1117
QY 348 LAGVAVSKLOKMATKNTINDSATKA--AVSQLSNLVGSVGFAGWTTAGLATPBAVKAE 406
DB 1118 ITANMALTQEKAAAIKQVDEBAKQAQAAVDAKSRADVDRAQDQIGKTSVPAVQPRK 1177
QY 407 SFTQDRKVSSTASSTSYVADQTVKLAKTVMDSGEAISSTGASLSTVNLRHSAP- 466
DB 1178 LNMIAAVDAQAATDKKAVINNDT--LTQEKKAALIRKVDDEAKARQAINDATSNADVAAK 1236
QY 467 IEEGGISAFSRSETP 481
DB 1237 QAQGTQAIANNVQPT 1251

```

RESULT 6

D84325
 Hcr17 transducer [imported] - Halobacterium sp. NRC-1

C/Species: Halobacterium sp. NRC-1
 C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C/Accession: D84325
 R/Ng, W.V.; Kennedy, S.P.; Mahatras, G.G.; Berguiet, B.; Pan, M.; Shukla, H.D.; Laeky, J.; Leithauser, B.; Keller, K.; Cruz, R.; Dawson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabb, Jung, K.H.; Alam, M.; Freilich, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li
 A/Title: Genome sequence of Halobacterium species NRC-1.
 A/Reference number: A84160; MUID:20504483; PMID:11016950
 A/Accession: D84325
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-536 <STO>
 A/Cross-references: GB:AE004437; NID:g10581193; PIDN:AAG1968.1; GSPDB:GN00138
 C/Genetic: hcr17
 A/Gene: hcr17
 C/Superfamily: Halobacterium salinarum transducer protein hcr1

Query Match 5.9%; Score 141; DB 2; Length 536;

Best Local Similarity 23.2%; Pred. No. 0.17; Matches 119; Conservative 78; Mismatches 197; Indels 120; Gaps 23;

```

QY 12 GVAMESFRTASDASLASVRSVSTTSCRDLOAITDYILKHHVFAHRSFVIGSPDERDAA 71
DB 63 GPVMDAFST---SLAVLIGVAIVISSTGTISIVED-LKTSNRAHR-----RRKDAQAK 113

```

Qy 72 LAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATATASSAFENTPPAAASVL 131
 Db 114 RRADQRAQAEAKAQAQDAREEAFTLAQKQERAEADAQLSKKASEYSVMKKAAS-- 171
 Qy 132 QYMQPAINKGDLWLAFLPLKPLTLI--SGALSGAMDQVG-----TKWMDRAR----- 175
 Db 172 -----GD-----LTVRANSSSDSDMTVEGQAQFEMIAAFEMVARTREFAGD 214
 Qy 176 -----GDLHYLSTSPDKLHDAMAVSVKRRHSPALGRQVQVDMGIAVQTFSSALNVRTVLAPA 230
 Db 215 VADNSRDIHASVDEIEAASEQVAESVQEISAGTDRE-----HTRLADA 257
 Qy 231 LASRPSVQAGVDFGVSTAGGLVANAGFGRMLSVQSRDLRGGAFLVGMKQKPEKPAALUSE 290
 Db 258 TDELOSGCAVE-EVAASASQVAD-----VSDQAADRGRQS-----DRAAAATQ 302
 Qy 291 ETDMLDAYKAIS-ASYSGAALNAGKRMAGLPDVLVATD--GLKAVRSLSVATSLTKGLA 347
 Db 303 MTDIOSETEAVVSEANALAAEAQETRRVVDLINDI-TDQTHLLANASIEARADQDGDG 361
 Qy 348 LAGGVAGVSKLQWATKNITD-----SATKAAVSQLSNLVSGVGFAGMTTAGLA 397
 Db 362 FAVVAAEVKELAE-ETNEATDEIETLVNLLSSTQSVDDSEQMAATVS--SGTETVENA 418
 Qy 398 -----TDPVAKKAESFIQDKVKSTASTSTYSVADQTVKLAKTVKMSGEAISST 446
 Db 419 LTALBEIGHVTD--VNEAVQOIDDATDAQAAS-----ADTTVSMIEIEDIS-EQNSSE 470
 Qy 447 GASLRSTVNNLRHRSAPADIEEGGISAFSRSET 480
 Db 471 AASVSA--QQQTASLSVDVTDGVTRLVDRADT 501

RESULT 7
 H89960
 hypotheical protein SA1577 [imported] - Staphylococcus aureus (strain N315)
 C;Species: Staphylococcus aureus
 C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C;Accession: H89960
 R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A;Reference number: A89758; PMID:21311952; PMID:11418146
 A;Accession: H89960
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-2186 <KUR>
 A;Cross-references: GB:BA000018; PID:g13701552; PIDN:BA842845.1; GSPDB:GN00149
 A;Experimental source: strain N315
 C;Genetics:
 A;Gene: SA1577

Query Match 5.8%; Score 140.5; DB 2; Length 2186;
 Best Local Similarity 19.3%; Pred. No. 1.3;
 Matches 105; Conservative 94; Mismatches 203; Indels 143; Gaps 21;

Qy 20 TASDASL-----ASSSVRSVSTSCRDLOAITDYLKHVFAAHRFVSGVSPDERDAALAHN 75
 Db 930 TAAANTNLANANTADVQVKTNIAIGIQAITPATK-----VKTAKNA----- 972
 Qy 76 EQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATATASSAFENTPPAAASVLQYMQ 135
 Db 973 --IDKSAETQHTNIENNND--ATLEEQQAQQLLDQAVATAKQIN--AAADTNQEVA 1023
 Qy 136 PAINKGDMLATPLKPLTLISGALSGAMDQVGTQWMDRARGDLHYLSTSPKLDHAMAVS 195
 Db 1024 QAKDQGTQNIWVIQATQVKTDA--VNDKAREAITINATPGATREEKQEA 1075
 Qy 196 VKRHSFALGRQVDMGIAVQTFSSALNVRT-----VLAPALASRPSVQGA----- 240

Db 1076 INRVNTLKNEALTIGV-TSTTAMVNSIRDDAVNQICAVQPHVTYKKTATGVLNDLATAK 1134
 Qy 241 -----VDFGVSTAGGLVANAGFGRMLSVQSRDLRGGAFLVGMKQKPEKPAALUSE 275
 Db 1135 KQEIQNQNTNATTEBEKQVALNQVDQLATA---INNINQADTNAEVDQAQQLGTKAINAIQ 1191
 Qy 276 -----VLGMKDKPEKPAALSEETDMLDAYKAISASYSGAALNAGKRMV----- 318
 Db 1192 PNIVVKPAALQAINQHYNAKLAEINATPDATNDEKNAIIN--TLNQDROQAIESIKOANT 1249
 Qy 319 GLPLD-----VATDGLKAVRSLSVATSLTKNGLALAGGAGVSKLQKMA--TKNITDSATK 372
 Db 1250 NAEVDQAATVAENNIDAVQVVDVVKQAARDKIT-----AEVAKRIEAVKQTPNATDEKQ 1304
 Qy 373 AAVSOLSLSNVSGVGFAGMTTAGLATDPAVKKA--ESFIQDKVKSTASTSTYSV-----ADQ 427
 Db 1305 AAVNQINQL-----KQALNQINQNTQDVTDTTNNQAVNAIDNVEAEV 1348
 Qy 428 TVK-----LAKTVKMSGEAISSTGASLSTVNNLRHRSAPADIEEGGISAFSRSETP 481
 Db 1349 VIKPKATADIEKAVKEKQKQIDN-----SLDSTDNEKEVASQALAKEKEKALAAIDQAGTN 1404
 Qy 482 FQLRR 486
 Db 1405 SQVNO 1409

RESULT 8
 T12796
 probable transglycosylase - Bacillus subtilis phage SPBC2
 C;Species: Bacillus subtilis phage SPBC2
 C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999
 C;Accession: T12796; A69911
 R;Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.
 submitted to the EMBL Data Library, August 1997
 A;Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 prophage
 A;Reference number: Z17583
 A;Accession: T12796
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-2285 <LAZ>
 A;Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025510; PIDN:AAC13005.1
 R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
 A.; Ehrlich, S.D.; Emmerston, P.T.; Etian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
 A;Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A;Reference number: A69580; PMID:98044033; PMID:9384377
 A;Accession: A69911
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-2285 <KUN>
 A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14053.1; PID:e1183582;
 A;Experimental source: strain 168
 C;Genetics:
 A;Gene: yomI

Query Match 5.8%; Score 140; DB 2; Length 2285;
 Best Local Similarity 20.5%; Pred. No. 1.5;
 Matches 128; Conservative 80; Mismatches 224; Indels 192; Gaps 26;

Qy 15 MESFRTASDASLASSSVRSVSTSCRDLOAITDYLKHVFAAHRFV-IGSPDERDAALA 73

```

Db 251 IELYORQAVNQNLNTRVSSSMGSSNRQAVODYLN-----AVNSLNVSTGSSNNIRSOIQS 306
Qy 74 HNEQIDAL---VETRANRLYSEGETPATIAET-----AKA 106
Db 307 LNNQFRELASNAQTAANQASSFG---AELTQTFKSMSTYLLISGLFYGAISGLKEMVSOA 363
Qy 107 EKRPDLATTSSAFENTPPFAASVLYQMPAINKGWMLATPLKPLTPLISGALSGAMDQV 166
Db 364 IETDTMTNIRRVNNEPDYKNEEL---QESIDLGDTLSKIKITDILQMTGDFGRMGFDES 420
Qy 167 GTKMPDRARDGLHYLS-TSPDKLHDAMAVSKRHPALGRQV-----VDMGIAVQTF 217
Db 421 ELSTLTKTAQVQLQNVDELTDVNTLTAMLPNTIANDSISLADKLNEVDNNVAVTTL 480
Qy 218 SALNVVETVLAPLASPSVQAVDPGVSTA-----GGLVAN-----AGFGDRLMS 263
Db 481 DLANSIRK---AGSTASTFVGLNDLIGYTTAIASTRESGNIVGNLSKTIPEARIGNQSS 538
Qy 264 VQSRDQLRGAFVLGKDKKPKAALSF---ETDML-DAYKAISASYSG----- 308
Db 539 IKALEIGISVKTAGEAKSASDLISEVAGKMDTLSDAQKNTSIGVAGIYOLSRFNAM 598
Qy 309 ---AALNAGK----- 315
Db 599 NNSFIQNAKTAANSTGSAWSEQKTKADSLQARVNLQNNFTFEPAIADAFISDGLIE 658
Qy 316 ---RMAGLPLDVAATDGLKAV---RSLVSATS-----LTKNGLALAGG-YAGVSKL--QKMA 362
Db 659 FTQAGSLNASTGVIKSVGLFPLLAAVSTATLLSKNRTLASLILGTRMAGQETTLA 718
Qy 363 TKNITSATGAASQ-----LSNLVGSVGVFAGTTAGLATDP-VKKA----- 405
Db 719 TAGLEAGMTRAAVASRVLKALTALRGLLVSTLVGAFAALGALBSLISSEFAKAKADDFE 778
Qy 406 -----ESFIIQ-----DKYKSTASTSTSYVADQ---TVGLAKT---VK 436
Db 779 QSQQTVEATTTMKDSTDKLIQYKELQYKESRSLTSDGEYLOYQVLAOTFPALVK 838
Qy 437 ---DMSGEAISSTGASLRSTVNNLR 458
Db 839 GYDSQGNALIKTKNELKAIENIK 862

RESULT 9
probable tail fiber protein GP37 - Escherichia coli (strain K-12)
C.Species: Escherichia coli
C.Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C.Accession: G64887; J09189
R.Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
Science 277, 1453-1462, 1997
A.Title: The complete genome sequence of Escherichia coli K-12.
A.Reference number: A64720; PMID:97426617; PMID:9278503
A.Accession: G64887
A.Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-1122 <BLAT>
A.Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AACT4454.1; PID:g1787636;
A.Experimental source: strain K-12, substrain MG1655
R.Albe, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.; Kasai, H.
; Motomura, K.; Nakade, S.; Nakamura, Y.; Nishimoto, H.; Nishio, Y.; Oshima, T.; Saito,
moto, Y.; Horiuchi, T.
DNA Res. 3, 363-377, 1996
A.Title: A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the
A.Reference number: Z16603; PMID:9751537; PMID:9097039
A.Accession: T09189
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 3-1122 <AIB>
A.Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AACT4454.1; PID:g1787636;

Query Match 5.7%; Score 138; DB 2; Length 1122;

```

```

Best Local Similarity 21.1%; Pred. No. 0.76;
Matches 106; Conservative 65; Mismatches 221; Indels 110; Gaps 17;

Qy 3 INOAAQPPGVAMSPRTASDASSVRSVSTTSCRDLOAITDYLKHHVFAHRSVI 62
Db 112 VEEVARNASAVQNTAAKKSASDASTSAREATHA----- 148
Qy 63 GSPDERDAALHNEQIDALVETRANRLYSEG--ETPATIA-ETFAKAEKDRLATASSA 119
Db 149 ---DAADSAARAASSTAQAASSAQSASSAGTASTATEASKAAAAESSKSAATASGA 205
Qy 120 FENTPPAASVLYQMPAINKGWMLATPLKPLTPLISGALSGAMDQVGTMMRAGDLH 179
Db 206 AKTSETNASASIQS-----AATSASTATTKASEATSRDAAAK--EAAKSET 253
Qy 180 YLSTSPD-----KLHDAMAVSKRHPALGRQV-VDMGIAVQTFPALNVVRT 225
Db 254 MASSSSAASASTAAGNSAKKAKTSETNARSETTAAAGSAAASAKTAAASASAST 313
Qy 226 VLAPALASPSVQAVDPGVSTAGLVANAGFGRMLVQSRDQLRGAFVLGKDKKPEK 285
Db 314 SAGQASASATPAACKSAESAASASTATTKAGATBQASAAAS-----ASAAKTSETN 366
Qy 286 ALSEETDMLDAYKAIKSASYSGALNAGKRNAGLPLDVAATDGLKAVRSLVATSLTNG 345
Db 367 AKASETS---AESKTAASASASSAASASAS--ASAKDEATROASAKS--SATTASTKA 420
Qy 346 LALAGVAGVSKLQKVAATNTDSTATKAASVQSLNSLVGSVGVFAGTTAGLATDPVKKA 405
Db 421 TENAG-----SATTAOSKSTA-----ESATRAETAKRA 451
Qy 406 ESFIIQKVSASTSTSYVADQTVKLAKTVKMSGEAISSTGSLRSTVNNLRHRSAP- 464
Db 452 EDIASAVALEDASTTKGI---VQLSATNSTRS-ETLAATPAVAVSAVDNAKRLQKQ 506
Qy 465 --ADIEEG-----ISAFSRSE 479
Db 507 NGADIDPKGCFNNINAVSKTD 528

RESULT 10
B98241
hypothetical protein AGR_L_1764 [imported] - Agrobacterium tumefaciens (strain C58, Cerc
C.Species: Agrobacterium tumefaciens
C.Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C.Accession: B98241
R.Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A.Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A.Reference number: A97359; PMID:11743194
A.Accession: B98241
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-1731 <KTR>
A.Cross-references: GB:AE007870; PIDN:AAK89452.1; PID:g15159316; GSPDB:GN00170
A.Genetics:
A:Gene: AGR_L_1764
A:Map position: linear chromosome

Query Match 5.7%; Score 138; DB 2; Length 1731;
Best Local Similarity 21.4%; Pred. No. 1.4;
Matches 116; Conservative 76; Mismatches 197; Indels 152; Gaps 23;

Qy 1 MHINQAQPPGVAMSPRTASDASSVRSVSTTSCRDLOAITDYLKHHVFAHRS 60
Db 746 MHIAELGTED---LAKFTKQIAIAISSAIAGLSA-----EALASLTQAQIALNTOS 795
Qy 61 VIGSPDERDAALHNEQIDALVETRANRL-----YSEGETPATIAETPA- 104
Db 796 INALSTRQIALALT-TQVVALTSTQVNALTSKQIALASTDDIATFSTKQIAAIAINSALIG 854
Qy 105 -KAEKDRLATASSAFENTPPFAASVLYQMPAINKGWMLAT-PLKPLTPLISGALSGA 162

```

Db 855 LSAETIASLATGQIAALNVRSTVQIVA-----LTTAQVEALTTVQVGLSST 906
 Qy 163 MDQVGTMMDRARGDLHYLSTSPDKLHDAMAVSVKRRHSPALGRQVV----- 208
 Db 907 -----QLAVLST-----DDIATFSTRDMAALGSSAIAGLSKDVTASLTTA 946
 Qy 209 DMG-----IAVQTFSSALNVRTVLAPALASRP-----SVQAVDFGVSTA 248
 Db 947 QIGALSMAGISGLSTGQIAALTDQNLVNTQIAAALTSKQVAAPFDVSDITALSQTGIAA 1006
 Qy 249 GGLVANAGFGDRMLSVQSRDQLRGCAFLGKMKPEKPAALSEE-----TWLDAYKAIK 302
 Db 1007 LSAAGAAGLTTDQIAAALSTQDV--GAMTSGQ-----IAALSAKQIAALGTDIATFTSGD 1059
 Qy 249 GGLVANAGFGDRMLSVQSRDQLRGCAFLGKMKPEKPAALSEE-----TWLDAYKAIK 302
 Db 1007 LSAAGAAGLTTDQIAAALSTQDV--GAMTSGQ-----IAALSAKQIAALGTDIATFTSGD 1059
 Qy 303 SASYSGAALNAGKRMAGLPLDVATDGLKAVRSLSVATSLSKNGLALAGGY----- 352
 Db 1060 IAAUSSNA-----VAGLSRDTVASLTTAQIAALSSAGISGLGTGQIAGLTSEQVNVLTN 1113
 Qy 353 AGVSKL--QKMATKNITDSATKAASVQSLNLSVGVGVAGTWTAGLATDP-----AVKAE 406
 Db 1114 AQISALTSKQVAALVETDIASLSA-AQIA-AIGAAGV-----AGLTTDQIAALSISQVE 1165
 Qy 407 SFIQDKVKSTASTTSYVADQTVKL-----AKTVKMSGEAISSTGA 448
 Db 1166 ALTSQAIAALNSKQIAALSAADDLAIFTTAEAAIGSGAISGLPASTIASLTTAQIAALGA 1225
 Qy 449 S 449
 Db 1226 A 1226
 RESULT 11
 AB3045
 ice nucleation protein homolog [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C;Species: Agrobacterium tumefaciens
 C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C;Accession: AB3045
 R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A;Reference number: AB2577; PMID:11743193
 A;Accession: AB3045
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1731 <KUR>
 A;Cross-references: GB:AB008689; PIDN:AAL44776.1; PID:g17742414; GSPDB:GN00187
 A;Experimental source: strain C58 (Dupont)
 C;Genetics:
 A;Gene: ina
 A;Map position: linear chromosome
 Query Match 5.7%; Score 138; DB 2; Length 1731;
 Best Local Similarity 21.4%; Pred. No. 1.4;
 Matches 116; Conservative 76; Mismatches 197; Indels 152; Gaps 23;
 Qy 1 MHINSAQOPPGVAMESFRTASDASLASSSVRSVSTTSCRDLOALTDVYLKHVFAHRFSV 50
 Db 746 MHIAELGTED----LAKFTTKDIAISSAIAGLSA-----EATASLTTAQIAALNTQS 795
 Qy 61 VIGSPDERDAALAHNEQIDALVETANRL-----YSEGETPATIAETFA- 104
 Db 796 INALSTAQIAALT-TAQVEALSTQVNTALTSKQIAALSTDDIATSTKDIAINSDAIG 854
 Qy 105 -KAEDFRLIATASSAFENTPFAASVLYQMOPAINKGDWLAT-PLKPLTPLISGALSA 162
 Db 855 LSAETIASLATGQIAALNVRSTVQIVA-----LTTAQVEALTTVQVGLSST 906
 Qy 163 MDQVGTMMDRARGDLHYLSTSPDKLHDAMAVSVKRRHSPALGRQVV----- 208

Db 907 -----QLAVLST-----DDIATFSTRDMAALGSSAIAGLSKDVTASLTTA 946
 Qy 209 DMG-----IAVQTFSSALNVRTVLAPALASRP-----SVQAVDFGVSTA 248
 Db 947 QIGALSMAGISGLSTGQIAALTDQNLVNTQIAAALTSKQVAAPFDVSDITALSQTGIAA 1006
 Qy 249 GGLVANAGFGDRMLSVQSRDQLRGCAFLGKMKPEKPAALSEE-----TWLDAYKAIK 302
 Db 1007 LSAAGAAGLTTDQIAAALSTQDV--GAMTSGQ-----IAALSAKQIAALGTDIATFTSGD 1059
 Qy 303 SASYSGAALNAGKRMAGLPLDVATDGLKAVRSLSVATSLSKNGLALAGGY----- 352
 Db 1060 IAAUSSNA-----VAGLSRDTVASLTTAQIAALSSAGISGLGTGQIAGLTSEQVNVLTN 1113
 Qy 353 AGVSKL--QKMATKNITDSATKAASVQSLNLSVGVGVAGTWTAGLATDP-----AVKAE 406
 Db 1114 AQISALTSKQVAALVETDIASLSA-AQIA-AIGAAGV-----AGLTTDQIAALSISQVE 1165
 Qy 407 SFIQDKVKSTASTTSYVADQTVKL-----AKTVKMSGEAISSTGA 448
 Db 1166 ALTSQAIAALNSKQIAALSAADDLAIFTTAEAAIGSGAISGLPASTIASLTTAQIAALGA 1225
 Qy 449 S 449
 Db 1226 A 1226
 RESULT 12
 B90835
 probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain RIM
 C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C;Accession: B90835
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
 A;Reference number: A99629; PMID:11156231; PMID:11258796
 A;Accession: B90835
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-971 <HAY>
 A;Cross-references: GB:BA000007; PIDN:BA035073.1; PID:g13361114; GSPDB:GN00154
 A;Experimental source: strain O157:H7, substrain RIMD 0509952
 C;Genetics:
 A;Gene: ECs1650
 Query Match 5.6%; Score 135; DB 2; Length 971;
 Best Local Similarity 22.0%; Pred. No. 0.97; 238; Indels 116; Gaps 20;
 Matches 116; Conservative 57; Mismatches 238; Indels 116; Gaps 20;
 Qy 3 INQSAQOPPGVAMESFRTASDASLASSSVRSVSTTSCRDLOALTDVYLKHVFAHRFSVI 62
 Db 110 VEEVARNASAVAQNTAAAKKSADASTSAREAT----- 143
 Qy 63 GSPERDAALAHNEQIDALVETANRLYSEG--ETPATIAETFAKAEKFDRLIATASSAF 120
 Db 144 HATDAADSARAASSTASAGQAASASQSSAGTASTKATKTEASAAAESSKSAATSA 203
 Qy 121 ENTFFAASVLYQMOPAINKGDWLATPLKPLTPLISGALSGAMDQVGTMMDRARGDLHY 180
 Db 204 AKTSETNAVSQ--QSA-----ATSASTATTKASEAASARDASAKSEAKS--SETSA 253
 Qy 181 LSTSPDKLHDAMAV--SVK-----RHSPALGRQVDDMGIAVQTFPSAL--NVVTVL 227
 Db 254 ASSASSAASATAGNSAKAKTSETNAKSSETAEQSASAAAGSKTAAALSASAASSTA 313
 Qy 228 APALASRPSVQGVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGCAFLGKMKPEKPA 287
 Db 314 QQASASATAAGKSAESAASSASTATTKAGEATEQASA-----AASSASAAKTSETNAK 366
 Qy 288 LSEETDMLDAYKAISASYSGAALNAGKRMAGLPLDVATDGLKAVRSLSVATSLSKNGLA 347

```

Db 367 ASBTS--AESSTKAAASASASASAS--ASASKEATQASAAKS--SATASTVATE 420
Qy 348 LAGYAGVSKLQ-----KMATKNITD-----SATKAIVSGLNVSIGV 387
Db 421 AAGSATTAAGSKTASBATAETRAETAKRADIASAVALEDASTTKKGIIVQSATNS--- 477
Qy 388 FAGMTAGLATDP-AVKKASFIDKVKSTASSTTSYVADQTVLAKTVKMGSEAISET 446
Db 478 ----TSBSLATPKAVKAAVELANGK-----YTAQDATTAQKGIIVQLS-NATNST 522
Qy 447 GASLRSTVNNLR-----HRSAPEDIEEGGT-----SAPSRSET 480
Db 523 SEMLAATPKSVKAAVDLANGKYTAQDATTAQKGIIVQSATNSASET 569

RESULT 13
C85693
probable membrane protein of prophage CP-933X Z1918 [imported] - Escherichia coli (strain
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C/Accession: C85693
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Pocamouns, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: AB5480; UID:21074935; PMID:11206551
A/Accession: C85693
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-973 <STO>
A/Cross-references: GB:AE005174; NID:G12514847; PIDN:AA656007.1; GSPDB:GN00145; UMGF:219
A/Experimental source: strain O157:H7, substrain EDL933
C/Genetics:
A/Genes: Z1918

Query Match 5.6%; Score 135; DB 2; Length 973;
Best Local Similarity 22.0%; Pred. No. 0.97;
Matches 116; Conservative 57; Mismatches 238; Indels 116; Gaps 20;

Qy 3 INSAOQPPVAMSEFTTASDASASSSVSVSTSCRDLOATIDYKHNFAHRSVI 62
Db 112 VEEVARNASVAQNTAAKKSASDASTAREBAT----- 145
Qy 63 GSPDERDALAHNEQIDALVETRAMRLYSEG--ETPATIETPAKKEPRLATTASSAF 120
Db 146 HATDADDSAAAEASTSAGQAAASASASAGTAKTEASKAAAESKSAATAGCA 205
Qy 121 ENTFFAAASVLYQWQPAINKGDMLATPLKPLTPLISGALSGAMDQVTKMDRAGDLY 180
Db 206 AKTSETNAASVQ--QSA-----ATGASTATTAKSEASARDAASAKS--SETSA 255
Qy 181 LSTSPDLHDMAMV--SVK-----HSPALGROVDKGIIVQFSL--NVATVL 227
Db 256 ASSASASATTAAGNAAKAKTSETNAESQASASAAAGSKTPAALISAAASTSA 315
Qy 228 APALASPSVQAGVDFGVSTAGGLVANAGFEDRMLSVQSRDQLEGAFVGMKKEPKAA 287
Db 316 GQASASATTAAGKASASASASTATTAKGATEQASA-----AASASAAKTSSETNAK 368
Qy 288 LSEETDMLADYKAIKSASYSGALINAKRMAGLEPLDVATDGLKAVRSLSATSLTKXGLA 347
Db 369 ASETT--AESSTKAAASASASASASAS--ASASKDEATQASAAKS--SATTAATATE 422
Qy 348 LAGYAGVSKLQ-----KMATKNITD-----SATKAIVSGLNVSIGV 387
Db 423 AAGSATTAAGSKTASBATAETRAETAKRADIASAVALEDASTTKKGIIVQSATNS--- 479
Qy 388 FAGMTAGLATDP-AVKKASFIDKVKSTASSTTSYVADQTVLAKTVKMGSEAISET 446
Db 480 ----TSBSLATPKAVKAAVELANGK-----YTAQDATTAQKGIIVQLS-NATNST 524
Qy 447 GASLRSTVNNLR-----HRSAPEDIEEGGT-----SAPSRSET 480

```

```

Db 525 SEMLAATPKSVKAAVDLANGKYTAQDATTAQKGIIVQSATNSASET 571

RESULT 14
AC0304
probable hemolysin YPO2490 [imported] - Yersinia pestis (strain CO92)
C/Species: Yersinia pestis
C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C/Accession: AC0304
R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B
deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davis, R.M.; Dougan, G.;
Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Bartell,
Nature 413, 523-527, 2001
A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A/Reference number: AB0001; UID:21470413; PMID:11585360
A/Accession: AC0304
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-2535 <KOR>
A/Cross-references: GB:AL590842; PIDN:CA091295.1; PID:G15980484; GSPDB:GN00175
C/Genetics:
A/Genes: YPO2490

Query Match 5.6%; Score 134; DB 2; Length 2535;
Best Local Similarity 23.0%; Pred. No. 4.3;
Matches 84; Conservative 66; Mismatches 168; Indels 48; Gaps 16;

Qy 126 AAASVLYQWQPAINKGDMLATPLKPLTPLISGALSGAMDQVTKMDRAGDLYHISTSP 165
Db 1383 ATSEHNRHLSSEVNVGALTAANDNLT-AVGANINATLEVOQNTS-----LSAAT 1433
Qy 186 DKHDMAMAVSVKGRSPALGRQVNDMGIAVQTFPALNVVTVTLPALASRPSVQAGVDFGV 245
Db 1434 DSHHTVGESSKRTSSV--NLVDFTLLSGQLNATGDINDQAAQDITLRS-----AV 1484
Qy 246 STAGGLVANAGFEDRMLSVGS--RDQLRGAFVLMKKEPKALSEETDMLADYKAIKS 303
Db 1485 QTDGALLTAAG--GDVLTITQTEQHDQRNHT--GLSKGIASSTLRTFEDSLTGLAVGS 1540
Qy 304 ASYSGAALNAGKRMAGLEPLDVATD---GLKAVRSLSVATS-----LTKNGIALA 349
Db 1541 MLGASGIDVSGKNIAMVGSNVVADODISLRAQENITVGTAAQSESSSHLFEOKKSGLMST 1600
Qy 350 GGYAGVSKLQKMATKNITDSTATAAVSQSLNVSIGVDFAGMTAGLATDPAYKASFI 409
Db 1601 GG-IGVT-VGSSSTK--MTDSG--QGISVSGTIGSV--LGNVSMRAGEDLRVQGAEVLA 1652
Qy 410 QDKVKSTASSTTSYVADQTVKLAITYD--MSGEAISSTGASLSRSTVNNLRHRSAPEDIE 468
Db 1653 GKOINLTGNKVSILAEENQLTQSHTEQKSGTLTALSGA--VGSANVTAVTTAKAASES 1711
Qy 469 EGGISA 474
Db 1712 SGRLGA 1717

RESULT 15
C87587
hypothetical protein CC2730 [imported] - Caulobacter crescentus
C/Species: Caulobacter crescentus
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: C87587
R/Merlan, W.C.; Feilblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heideberg, J
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Grimm, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of Caulobacter crescentus.
A/Reference number: AB7249; UID:21173698; PMID:11259647
A/Accession: C87587
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-654 <STO>

```

A;Cross-references: GB:AE005673; NID:g13424321; PIDN:AAK24695.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2730

```
Query Match      5.5%; Score 132.5; DB 2; Length 654;
Best Local Similarity 24.4%; Pred. No. 0.81;
Matches 116; Conservative 67; Mismatches 162; Indels 131; Gaps 28;

Qy 61 VIGSPDERDAALAHNEQIDALVETANRLYSEGETPATIAET-----FAK--AEKFDRLAT 114
Db 3 VVGS-----QNTALSSLSPTRIAL-----SAATLSEAPASEFAKLTAAQVAAISI 48

Qy 115 TASSAFENTPRAASVLYQMPAINKGDWLATPLKPLT-----PLISGALSGAMDQVGTGM 170
Db 49 TAI SALTSEQFAAFDVRQ-----VRALTS AQLPRLSAAISFSPKQLAT-- 91

Qy 171 MDRARGDLHYLSTSPDKLHDAMAVSVKRRHSPALGR-QVVDMGIAVOT-FSALNV----- 222
Db 92 -----LSQACTQALQHTQVNAISI-----TALGEMRTTQLAVLSATAPAAISIDQVQSL 140

Qy 223 ----VRTVLAPALASRPVQGVAVDFGVSTAGGLVA--NAGFGDRMLSVQSRDQLRGGAFF 276
Db 141 SLAQVRSLSAPQVRALPAADIA-EPSLEQVGALTALQLAVLSPTQVSALSSEQ----- 192

Qy 277 LGMKDKEPKAALSEET-----DWLDAYKAIKSASYSGAALNA--GKRMAGL 320
Db 193 IGALSRTOFSALSAATVRALSPTLALTTADHWRAATTIQAOMSSLSGQDKALGETRLSSL 252

Qy 321 PLDVATDGLKAVRSLSVTSLSLTNGKLALAGG-----YAGVSKLQKMATKNITDSATKAIV 375
Db 253 -----SASQVRAL-SATEVGRAGPALLDALTPTQLAGLGNH-TAALTVTYDIALGSA- 302

Qy 376 SOLSNLVGSVGVFAGWTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTVKL-AKT 434
Db 303 SQIAAL--DPAAVAGLSATGLSA-----LSASQVAALSASQVAVRLRANQ 344

Qy 435 VKDMSGEAISS-TGASLRS-TVNNLRH-RSAPEADIEGGISAFSESETPFQLREL 487
Db 345 LQGLNGEDFSEFTRAOLKAFATAQIRELPFARLAEISDEDFGEF----TATQIRAL 396
```

Search completed: January 2, 2003, 15:21:21
Job time : 26.0175 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:17:36 ; Search time 13.5139 seconds
(without alignments)
1060.315 Million cell updates/sec

Title: US-09-825-414-66

Sequence: 1 MHINSAQPPGVAMESFRT.....EEGIGAFSRSTPQLRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/aa/5A_COMB.pep:*
4: /cgn2_6/prodata/1/aa/5B_COMB.pep:*
5: /cgn2_6/prodata/1/aa/PCITUS_COMB.pep:*
6: /cgn2_6/prodata/1/aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146.5	6.1	2123	4	US-08-968-685A-10
2	140	5.8	2285	4	US-09-108-375-2
3	132	5.5	1026	4	US-08-614-377A-7
4	132	5.5	1026	4	US-09-142-648B-7
5	130	5.4	1026	4	US-08-194-290-7
6	125.5	5.2	1232	2	US-08-682-517-15
7	125.5	5.2	1232	2	US-08-682-517-9
8	123	5.1	688	3	US-09-141-047-8
9	122.5	5.1	1040	4	US-09-134-001C-5365
10	121.5	5.0	1068	4	US-09-085-199B-11
11	117	4.9	682	1	US-08-441-139-2
12	116	4.8	488	4	US-08-809-326A-1
13	116	4.8	649	4	US-08-809-326A-15
14	116	4.8	2887	4	US-08-462-467B-8
15	115.5	4.8	401	2	US-08-591-079-6
16	115	4.8	679	1	US-08-441-139-5
17	114.5	4.8	748	4	US-09-066-046-4
18	110.5	4.6	2137	4	US-09-134-001C-4463
19	110	4.6	2089	1	US-08-418-893D-23
20	110	4.6	2089	1	US-08-418-893D-24
21	109.5	4.5	1618	4	US-08-462-467B-4
22	109.5	4.5	2887	4	US-08-462-467B-2
23	108.5	4.5	1002	4	US-09-268-347-24
24	108	4.5	2314	4	US-09-268-347-49
25	108	4.5	5067	4	US-09-144-085-1
26	107.5	4.5	676	4	US-09-085-199B-7
27	107.5	4.5	1140	4	US-09-220-081-2

28	107.5	4.5	1140	4	US-09-677-575-2	Sequence 2, Appl
29	107.5	4.5	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
30	107	4.4	1004	4	US-09-268-347-30	Sequence 30, Appl
31	107	4.4	7257	3	US-09-335-409-5	Sequence 5, Appl
32	107	4.4	7257	4	US-09-568-102-5	Sequence 5, Appl
33	107	4.4	7257	4	US-09-567-969-5	Sequence 5, Appl
34	107	4.4	7257	4	US-09-568-480-5	Sequence 5, Appl
35	107	4.4	7257	4	US-09-568-486-5	Sequence 5, Appl
36	107	4.4	7257	4	US-09-568-472-5	Sequence 5, Appl
37	107	4.4	7257	4	US-09-567-899-5	Sequence 5, Appl
38	106	4.4	1833	4	US-08-621-944A-4	Sequence 4, Appl
39	106	4.4	1833	4	US-08-945-567D-4	Sequence 4, Appl
40	106	4.4	1992	4	US-08-621-944A-3	Sequence 3, Appl
41	106	4.4	1992	4	US-08-945-567D-3	Sequence 3, Appl
42	104.5	4.3	367	3	US-09-141-047-10	Sequence 10, Appl
43	104	4.3	547	2	US-08-467-822-35	Sequence 35, Appl
44	104	4.3	547	4	US-08-432-697-35	Sequence 35, Appl
45	104	4.3	547	4	US-08-466-248-35	Sequence 35, Appl

ALIGNMENTS

```
RESULT 1
US-08-968-685A-10
; Sequence 10, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: TUCKER, KENNETH
; APPLICANT: PLOSILA, LAURA
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,685A
; FILING DATE: NO. 6214981ember 12, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ. ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2123 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-968-685A-10

Query Match 6.1%; Score 146.5; DB 4; Length 2123;
Best Local Similarity 22.2%; Pred. No. 0.00012;
Matches 111; Conservative 69; Mismatches 181; Indels 139; Gaps 23;
CY 32 RSVSTTSCRLQAITDYKIHVFAHRFSYTGSPDERDALANEOQDALVER--ANR 88
DB 377 KTLITIGAGTSAITD-----HNIGVQNGDGKVKQLAETLTSLKMTTENTLTANE 427
```

Qy 89 LYSBETPATIAETPAKAEKFORLATTASSAFENTPPAAASVLYQMOPAINKGWDLATPL 148
Db 428 KVTVGKT-----RL-TTKIGFTND-----MNGIDBSKPYLDKDTGHIHAGG 467
Qy 149 KPLTPLISGAL-----SGAMDQVGTQKMDRARGDLHLVSTSPDKLHDA 191
Db 468 QKITKLTAGVDDDDAATYGQLKKVQNTAESALQFTTVKKVDKNGND-----ANDS 517
Qy 192 MAVSKRHPALGRQVDM-----GIAVQTFSAALNVVTVLAPALASRPSVQGAADFV 245
Db 518 KIITVGNKPKDGTQVNTLKLKENGVDVTT-----ETNGTVTFGL 558
Qy 246 STAGGL-VANAGFGDRLMLSVQ-----SRDQLRGAGFVLGMKD---KEPKAALSEET-----DW 294
Db 559 NONNGLTGVNSTLNDGLSVKNTNSNKIQVADGITTFTDISNSKPGAGIENTTTRITRDG 618
Qy 295 LDAYKAISASVSGAALNAGKEMAGLPDVAFD---GLKAVRSLSVATSLSLTKNGLALAGG 351
Db 619 I-----GPANNTGSLDANKPLTPTGINAGGKELTNVQSALNPATNG-----660
Qy 352 YAGVSKLOMATKNTIDTSATKAASVQSLNLYSGVGFAGTWTAGLATDPAV-KKAESFIQ 410
Db 661 -GOLDFMRLSTANTEKSGSAATIKDLNLSQVPLTFAG-----DTGPNVTKKLGEILK 713
Qy 411 DK-VKSTASST-----SVDAQ-----TVKLAKTVKDMSGEASSTGASLSRSTVNNLHR 460
Db 714 VKGGKTTADDLTKNNIGVAVDSTONSLTVKLAKTSLDLDVNTKTLTASDKVTVDSGNN- 772
Qy 461 SAPEADIEGGISAFSRSSET 480
Db 773 ---TAKLQNGDLT-FSKONT 788

RESULT 2

US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCT
; CURRENT APPLICATION NUMBER: US/09/308,375
; CURRENT FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-308-375-2

Query Match 5.8%; Score 140; DB 4; Length 2285;
Best Local Similarity 20.5%; Pred. No. 0.00058;
Matches 128; Conservative 80; Mismatches 224; Indels 192; Gaps 26;

Qy 15 MESPTASDASLASSSVRSVSTSCRDQLQAITDYLKHVFAHRSFV-IGSPDERDAALA 73
Db 251 IELYORQVQVNVNLTNYGSMGSSNQAVQDYLN-----AVNSLVNVTGSGNNIRSQIS 306
Qy 74 HNEQIDAL---VETRANRLYSEGETPATIAETF-----AKA 106
Db 307 LNMQPRELASNAQTAANQASSFG---AELTQTFKSMSTYLSGSLFYGAISGLKEMVQA 363
Qy 107 EKDFRLATTASAFENTPFAASVLYQMOPAINKGWDLATPLKPLTPLISGALSAMQOV 166
Db 364 IEIDTMTNIRRVNNEPYKYNNELL---QESIDLGTLSNKTIDTLQMTGDFGRMGFDES 420
Qy 167 GTKMDRARGDLHLYS-TSPDKLHDAMAVSVKRHSPALGRGV-----VDMGIATQTF 217
Db 421 ELSTLTAKVQLVQVSDLTPTDDTVNTLTAAMLNFNIAANDSISIADKLNVDNNVAVTTL 480

Qy 218 SALNVTVTLAPALASRPSVQGAADFVGSTA-----GGLVAN-----AGFGDRMLS 263
Db 481 DLANSIRK--AGSTASTFGVELNDLIGYTTAIAASTTTRESGNIVGNSLKTFIFARIGNQSS 538
Qy 264 VOSRDQLRGAGFVLGMKDKEPKAALSE---ETDWL-DAYKAISASVSG-----308
Db 539 IKALEQIGISVKTAGGAKSASDLISEVAGKWDTLSDAQKNTSIGVAGIYQLSRFNMM 598
Qy 309 ---AALNAGK-----315
Db 599 NNFSAQNAAKTAANSTGSAWSEQKYADSLQARVKNLQNNFTFAAASDAFISDGLIE 658
Qy 316 --RWAGLPDVAVDGLKAV---RSLVSATS-----LTKNGLALAGG-YAGVSKL--QKMA 362
Db 659 FTQAAGSLLNASTGVIKSVGFPLPPLAAVSTATILLKSKNTRTLASSLILGTRAMGOETLA 718
Qy 363 TKNTIDTSATKAAVSQ-----LSNLVGSVGVFAGTWTAGLATDPA-VKKA-----405
Db 719 TAGLEAGMTAAVASRVLKTALRGLVSTLVGGFAALGMALESLSISFAEAKKAKKDDFE 778
Qy 406 -----ESFIQ-----DKVKSTASSTTSYVADQ---TVKLAKT---VK 436
Db 779 QSQQTNVAEITTKNDSTDKLIQQYKELQKVESRSLTSDBEQEYLVQVTTQQAQTFPALVK 838
Qy 437 --DMSGASISTGASLSRSTVNNLR 458
Db 839 GYDSQGNAILTKNKELEKAIENTK 862

RESULT 3

US-08-614-377A-7
; Sequence 7, Application US/08614377A
; Patent No. 5976864
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. 5976864ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF
; TITLE OF INVENTION: HETEROLOGOUS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson PC
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/614,377A
; FILING DATE: 12-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/194,290
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 435
; APPLICATION NUMBER: US 07/895,367
; FILING DATE: 09-JUNE-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsao, Y. Rocky
; REGISTRATION NUMBER: 34053
; REFERENCE/DOCKET NUMBER: 08106/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1026 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-614-377A-7

Query Match 5.5%; Score 132; DB 2; Length 1026;
 Best Local Similarity 21.0%; Pred. No. 0.00092;
 Matches 101; Conservative 75; Mismatches 166; Indels 140; Gaps 23;

QY 22 SDASLASSSVRSVSTSCDQLQITDYDKHVPAAHRSVIGSPDERDAALAHNEQIDAL 81
 DB 42 SDAALNTLTKLVNSTTAAVAIQ-----YQFTGVAPSAAG-----LDPL 81
 QY 82 VETRAN-----RLYSEGETPATIAETFAKAEK-----DLATTASAFENTPPAA--- 128
 DB 82 VDSITNTNDLNDAYYSK-----FAQENRFINFSINLATGAGAG--ATFAAAAYTG 129
 QY 129 -SVLYQWQPAINK--GDMLATPLKPLPLISGALSGAMPDQGTMMDRARGLHYLSTSP 185
 DB 130 VSTAQIVATAYDKIIGNAVAT-----AAGVDVAANAVALSRQANIDYLT-- 173
 QY 186 DKLDMAVSVKHSPPALGRQVDMGIAVQTFSAINVTVL--APALASRPSVQAVD 242
 DB 174 -----AFVRANTPPTAADIDLAVKALIGTILMAATVSGIGYATATAMINDLSD 225
 QY 243 FGYST--AGGL-----VANAGFGDRMLSVOG--RDQLRG-----GAFVLGMKD 281
 DB 226 GALSTDMAGVNLFTAYPSGVSGLSLTGTGDTLTGTANNDFVAGVAGATTLTVGD 285
 QY 282 KEPKALSEETDMLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLSVATSL 341
 DB 286 TISGAGTIDVLMVQA-----AAVTA-----LPTGVITISGIERM--NVTSGAAL 327
 QY 342 TKXGLALAGYAGVSKL-----QKMATKNIITSATRAAVSOLSNL--VGS 384
 DB 328 TLN---TSSGVTLTLTALNTNTSGAQTVTAGAGQNLATTTAAQAANNVAVADGRANVTVAS 384
 QY 385 VGVFAGTTTGLATDPAVKKAESFIQDKVKSSTTSYVADQTVKLAKTVKMSGEAIS 444
 DB 385 TGVTSGETTVG-----ANSASGTVSVANSSTTTTGALA--VTGGATVVAQTAGNAVN 438
 QY 445 ST 446
 DB 439 TT 440

RESULT 4

US-09-142-648B-7
 Sequence 7, Application US/09142648B
 Patent No. 6210948
 GENERAL INFORMATION:
 APPLICANT: Smit, John
 APPLICANT: Bingle, Wade H.
 APPLICANT: No. 6210948ellini, John F.
 TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULIC
 FILE REFERENCE: 08106/002002
 CURRENT APPLICATION NUMBER: US/09/142,648B
 PRIOR FILING DATE: 1999-03-30
 PRIOR APPLICATION NUMBER: PCT/CA97/00167
 PRIOR FILING DATE: 1997-03-10
 PRIOR APPLICATION NUMBER: US 07/614,377
 PRIOR FILING DATE: 1996-03-12
 NUMBER OF SEQ ID NOS: 12
 SOFTWARE: FASTSEQ for Windows Version 4.0
 SEQ ID NO 7
 LENGTH: 1026
 TYPE: PRT
 ORGANISM: Caulobacter crescentus
 US-09-142-648B-7

Query Match 5.5%; Score 132; DB 4; Length 1026;
 Best Local Similarity 21.0%; Pred. No. 0.00092;
 Matches 101; Conservative 75; Mismatches 166; Indels 140; Gaps 23;

QY 22 SDASLASSSVRSVSTSCDQLQITDYDKHVPAAHRSVIGSPDERDAALAHNEQIDAL 81
 DB 42 SDAALNTLTKLVNSTTAAVAIQ-----YQFTGVAPSAAG-----LDPL 81
 QY 82 VETRAN-----RLYSEGETPATIAETFAKAEK-----DLATTASAFENTPPAA--- 128
 DB 82 VDSITNTNDLNDAYYSK-----FAQENRFINFSINLATGAGAG--ATFAAAAYTG 129
 QY 129 -SVLYQWQPAINK--GDMLATPLKPLPLISGALSGAMPDQGTMMDRARGLHYLSTSP 185
 DB 130 VSTAQIVATAYDKIIGNAVAT-----AAGVDVAANAVALSRQANIDYLT-- 173
 QY 186 DKLDMAVSVKHSPPALGRQVDMGIAVQTFSAINVTVL--APALASRPSVQAVD 242
 DB 174 -----AFVRANTPPTAADIDLAVKALIGTILMAATVSGIGYATATAMINDLSD 225
 QY 243 FGYST--AGGL-----VANAGFGDRMLSVOG--RDQLRG-----GAFVLGMKD 281
 DB 226 GALSTDMAGVNLFTAYPSGVSGLSLTGTGDTLTGTANNDFVAGVAGATTLTVGD 285
 QY 282 KEPKALSEETDMLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLSVATSL 341
 DB 286 TISGAGTIDVLMVQA-----AAVTA-----LPTGVITISGIERM--NVTSGAAL 327
 QY 342 TKXGLALAGYAGVSKL-----QKMATKNIITSATRAAVSOLSNL--VGS 384
 DB 328 TLN---TSSGVTLTLTALNTNTSGAQTVTAGAGQNLATTTAAQAANNVAVADGRANVTVAS 384
 QY 385 VGVFAGTTTGLATDPAVKKAESFIQDKVKSSTTSYVADQTVKLAKTVKMSGEAIS 444
 DB 385 TGVTSGETTVG-----ANSASGTVSVANSSTTTTGALA--VTGGATVVAQTAGNAVN 438
 QY 445 ST 446
 DB 439 TT 440

RESULT 5

US-08-194-290-7
 Sequence 7, Application US/08194290
 Patent No. 5500353
 GENERAL INFORMATION:
 APPLICANT: Smit, John
 APPLICANT: Bingle, Wade H.
 TITLE OF INVENTION: Bacterial surface protein expression
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Shlesinger, Arkwright & Garvey
 STREET: 3000 South Eads Street
 CITY: Arlington
 STATE: Virginia
 COUNTRY: USA
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/194,290
 FILING DATE: 09-FEB-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Garvey, George A.
 REGISTRATION NUMBER: 17737
 REFERENCE/DOCKET NUMBER: 5946-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-684-5600
 TELEFAX: 703-836-5288

INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1026 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-194-290-7

Query Match 5.4%; Score 130; DB 1; Length 1026;
 Best Local Similarity 21.0%; Pred. No. 0.0014;
 Matches 101; Conservative 75; Mismatches 166; Indels 140; Gaps 23;

QY 22 SDASLASSVRSVTSCELDQAIIDYKHHVFAHRSVIGSPDERDAALAHNEQIDAL 81
 DB 42 SDAALITNLKLVNSTTAVAIQT-----YOFFTGVAPSAAG-----LDLF 81
 QY 82 VETRA-----RLYSEGETPATIAETFAKAEKF-----DRLATTASSAFENTPFAA--- 128
 DB 82 VDSITNTDNDAYISK-----FAQENRFNFINSINLATGAGAG--ATAFAAAVTG 129
 QY 129 -SVLYQMOPAINK--GDWLATPLKPLTBLISGALSGAMDQVGTQMDRARGDLHYLSTSP 185
 DB 130 VSYAQTVATVDKIIIGNAVAT-----AAGVDVAAVAFLSRQANIDYLT--- 173
 QY 186 DKLHDAMAVSKHSPALGRQVDMGIAVQTFSAALNVRTVL---APALASRPSVQGAVD 242
 DB 174 -----AFVRANTPFTAAADIDLAVKAALIGITILNAATVSGIGGYATATAAMINDLSD 225
 QY 243 FGVT--AGGL-----VANAGFGDRMLSVQS--RDQLRG-----GAFVLGMKD 281
 DB 226 GALSTDNAAGVNLFTAYPSSGVSGSTLSLTGTDTLTGTANNDFVAGEVAGATLTVD 285
 QY 282 KEPKAALSEETDMDLDAKIAKSAVSYSYGAALNAGKMGALPLDVAIDGLKAVRSLSVATSL 341
 DB 286 TLLSGAGTDVLMWQA-----AAVTA-----LPTGVTISGIETM-NVTSAAI 327
 QY 342 TKNGLALAGGAGVSKL-----QKWAFTNITDSATKAASVLSNL-VGS 384
 DB 328 TLN-----TSSGVTLTALNTNTSGAQTVTAGAGQNLTTATAQAANNVAVDGGANVTVAS 384
 QY 385 VGVFAGWTTAGLATDPVKKAESEFTQDKVKSTPASTTSVSDQTVKLTAKVDMSEGAIS 444
 DB 385 TGVTSGTITVG-----ANSAAGTSVSVANSSTTTTGIAA-VTGTAVTVAQTAGNAVN 438
 QY 445 ST 446
 DB 439 TT 440

RESULT 6
 US-08-682-517-15
 ; Sequence 15, Application US/08682517
 ; Patent No. 5874267
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Expression of surface layer proteins
 ; NUMBER OF SEQUENCES: 25
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/682,517
 ; FILING DATE:
 ; CLASSIFICATION:
 ; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1222 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-682-517-15

Query Match 5.2%; Score 125.5; DB 2; Length 1222;
 Best Local Similarity 20.3%; Pred. No. 0.0053;
 Matches 117; Conservative 73; Mismatches 219; Indels 167; Gaps 25;

QY 3 INQSAQQPPGVAME-----SFRITASDASLASSSVRSVSTT 37
 DB 675 VNQSGKVTGTSIKKATYTIYNTGANDIKVDNQVISPNSYTVTYEATLSST---GVIT 731
 QY 38 SCRDLQAITDYLKHHVFAHRSVIGSPDERDAALAHNEQIDALVET----- 84
 DB 732 PAKLEVTSDGKTAVKVIATGIANTDGDYAKTAKEATATATATNEVPNSVTGVATQ 791
 QY 85 -----RANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLYQ--- 133
 DB 792 FNTADSGSNSINWFAGKNPVKAGVSGKTYKY--FGANGNEVFGEAAWEAL-LTQYATE 848
 QY 134 -----MQPALNKGDWLATPLKPLTPL--ISGALS-----CAMD--- 164
 DB 849 GQKVTISYNDGDTVTFKVISAVNSSTEAIKPVAPTTPAAPTGGALTLTPAAGGLVDLTT 908
 QY 165 QVGTQMMDRARGDLHYLSTSPDKLHDAMAVSVK-RHSPALGRQVDMGIAVQTFSAALNV 223
 DB 909 ATNTLIGISLADADLNVSAITV---DTATVSLKDSANNSLSLTIVETGANTGVFAT--- 960
 QY 224 RTVLAPALASRPSVQGAVDGFGVSTAGGLVAN-----AGFGDRMLSVQSRDQLRG--- 273
 DB 961 -TVQAGTLLS-----LTAGTLTVTYADAKNAAGVAENITASVTLKTKTTGATS 1007
 QY 274 -AFVLGMKDKEPKAALSEETDMDLDAKIAKSAVSYSYGAAL--NAGKMGALPLDVAIDGL 329
 DB 1008 DTFQGVL---PSANTAEYTSKIAADYTFATGEGTLLDNAGAQVINA-----GK 1058
 QY 330 KAVRSLSVATSLTKNGLALAGGAGVSKLQKMAFTKNTD-----SATKAASVLSNLV- 382
 DB 1059 KGAQGVADAINATPAGTATVSG-----DKVVIKSATTVGVSEVEVTFSSVNOVLNAV 1111
 QY 383 -GSVGVFAGW-TTAGLATDPAVKAESFIQKVKVST-----ASSTYS 423
 DB 1112 NGDQVAVAGTAATKRAFTITTSLSVGEKVWIDGVETAVAFGTAPTANTFVVEESAANTLAS 1171
 QY 424 VADQTVKLAKTKVDMSE---AISSTGA--SLRSTV 454
 DB 1172 VADQAAANLAATIDTLNTADKFTASATGATTLTSTV 1207

RESULT 7
 US-08-682-517-9
 ; Sequence 9, Application US/08682517
 ; Patent No. 5874267
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Expression of surface layer proteins
 ; NUMBER OF SEQUENCES: 25
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/682,517
 ; FILING DATE:
 ; CLASSIFICATION:
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1252 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-682-517-9

Query Match 5.2%; Score 125.5; DB 2; Length 1252;
 Best Local Similarity 20.3%; Pred. No. 0.0055;

Matches 117; Conservative 73; Mismatches 219; Indels 167; Gaps 25;

QY 3 INSOAQQPFGVAME-----SFRISADSLASSVSRYSTT 37
 DB 705 VNOSGKVTGSIKAKAYTTTNTGANDIKVDNOVISPNRSITTYEATLST---GRVIT 761
 QY 38 SCRDQAITDYLRKHVFAHRFVSIGSPDERDALAHNEQIDALVER-----84
 DB 762 PAKLEVTSTVDGKTAVKVIATGIAVNTDGDVAFPAKEATATPTAINEVNSYGVATQ 821
 QY 85 -----RANRLVSEGETPATIETFAKAEKPRRLATTSSAENPFAASLYQY--- 133
 DB 822 FNTADSGNSNSTWFAKKNPCKYGVSGKTYKY--FGANGNEVGEAAEWL-LTQYATE 878
 QY 134 -----MOPAINKGDMLATPLKPLTPL--ISGALS-----GAMD--- 164
 DB 879 GQKVTISYVNDGTVTPTKVISAVNSSTEAIKPVAPPTPAFTGALTUTPAAGLVLT 938
 QY 165 QVGTQMDRARGDLHYLSTSPDKLHDMANVSVR-RHSPALGRQVDMGIAVQTFSAINV 223
 DB 939 AITNTGLSLDADLVNATVTV---DTATVSLKDSANNSLSLTIVETGANTGVFAT---- 990
 QY 224 RTVLAPALASRPVQGVAVDVGSTAGLVAN-----AGFGDMLSVQSRDQLRG--- 273
 DB 991 -TVQAGTLSS-----LTAGTLTVTYADAKNAAGVAENITASVTILKTTGATIS 1037
 QY 274 -AFVLGKMKDEPKAALSEETDMLDAYKAISASYSGAAL---NAGKMAGLPLDVTGDL 329
 DB 1038 DTFITQGLV---PSAATAEYTSKSIADYTPATGEGFTLINDNAGQVYNLA-----GK 1088
 QY 330 KAVRSLSVATSLTRNGALAGVAVSKLQPMATKNITD-----SATKAVSQLSNLY- 382
 DB 1089 KGAGGVADAINATPAGTAVTVSG-----DKVVIKSATTVGVSEVETPSVQVLANAV 1141
 QY 383 -GSVGVFAGM-TTAGLATDPAVKKAESEFIODKXST-----ASSTTSY 423
 DB 1142 NGKQOVAGTAATKATFTTTTLISVGEKVVIDGVEYTVAVAGTAPANTFVESANATLAS 1201
 QY 424 VADQTVLAKTVKDMGSE---AISSTGA--SLRSTV 454
 DB 1202 VADQANLAAITDILNADKFTASATGATITLTSTV 1237

RESULT 8
 US-09-141-047-8
 ; Sequence 8, Application US/09141047A
 ; Patient No. 6063085
 ; GENERAL INFORMATION:
 ; APPLICANT: Walker, David H.
 ; APPLICANT: Yu, Xue-Jie
 ; TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
 ; TITLE OF INVENTION: Protein Gene
 ; FILE REFERENCE: D6143
 ; CURRENT APPLICATION NUMBER: US/09/141,047A
 ; CURRENT FILING DATE: 1998-08-27
 ; NUMBER OF SEQ ID NOS: 15
 ; SEQ ID NO 8
 ; LENGTH: 688
 ; TYPE: PRF
 ; ORGANISM: Ehrlichia canis
 ; FEATURES:
 ; OTHER INFORMATION: Amino acid sequence of 120 kDa immunoreactive
 ; OTHER INFORMATION: protein.
 US-09-141-047-8

Query Match 5.1%; Score 123; DB 3; Length 688;
 Best Local Similarity 21.1%; Pred. No. 0.0036;
 Matches 114; Conservative 78; Mismatches 205; Indels 144; Gaps 26;

QY 8 QOPPGVAMESPRFASDASLASSV---RSVSTTS-----CRDIO-AITDYLKHHVF 54
 DB 105 ESTPEVAAEDLQPAVDSDIEHSSEVGEKVSKESTPEVKAEDLQPAVDSDVE--- 160

QY 55 AAHRFVSIVG-----SPDER-----DALAHNEQIDALVETRANRLYSEGET 95
 DB 161 --HSSSEVGEKVSKEENTPEVKAEDLQPAVDGSIEHS---SSEVGEKVSKEEST 215
 QY 96 PATIAETP-----AAAEKPRRLATTASSAFENTPFAASVLYQMPALNKGWLA 145
 DB 216 PEVKAEDLQPAVDSEVHSSEVGEKVSKE--ENTPEVKAEDLQ-----260
 QY 146 TPLKPLTPLISGALSGAMDQVGTQMDRARGDLHYLSTSPDKLHD--AMAVSKRSPA 202
 DB 261 -----PAVDGSEVHSSEVGVKVSKEE---STPEVKAEDLQPAVDSEVHSSE 309
 QY 203 LGRQVDMGIAVQTFSAINVTVTLAPALAS--RPSVQAVDGVSTAGLVANAGFGD 259
 DB 310 VGEKVSKEENT-----PEVKAEDLQPAVDGSEVHSSEVGEKVSKEE 357
 QY 260 RMLSVQSRD-----QLRGAFVLGKMKDE-PRKALSEETDMLDAYKAI-----KSASY 306
 DB 358 STEPEVKAEDLQPAVDSDIEHSSEVGEKVSKEESTPEVKAEDLQPAVDGSEVHSSE 417
 QY 307 SGALNAGKMGALP-----LDVATDG-----LKAVRSLSVATSLTRNGALAGYAGV 355
 DB 418 VGEKVSKEENTPEVKAEDLQPAVDGSEVHSSEVGEKVSKEENTPEVK-----A 472
 QY 356 SKLQPMATKNITDSATKAA--VSQLSNLYGVGVFAGMTAGATDPAYKKAESFIODKY 413
 DB 473 EDLQPAVDGSEVHSSEVGEKVSKEESTPEVKA--EDLQPAVDGSEVHSSEVGEKV 530
 QY 414 KSTN--SSTISYVA-DQTVLAKTVKDMG---GEAISSTASLSRSTVNNLRHSAPRADI 467
 DB 531 SETSKESTPEVKAEDLQPAVDGSEVHSSEVGEKVSKESTPEVKA-----KEESTPEVKA 580
 QY 468 E 468
 DB 581 E 581

RESULT 9
 US-09-134-001C-5365
 ; Sequence 5365, Application US/09134001C
 ; Patient No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 5365
 ; LENGTH: 1040
 ; TYPE: PRF
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-5365

Query Match 5.1%; Score 122.5; DB 4; Length 1040;
 Best Local Similarity 19.3%; Pred. No. 0.008;
 Matches 95; Conservative 71; Mismatches 177; Indels 149; Gaps 19;

QY 45 ITDYLRKHVFAHRFVSIGSPDERDALAHNEQIDALVETRANRLYSEGETPATIETFA 104
 DB 19 ITDF-QNIRKAKRMAKTEIIPDE-----IEQVQV-----A 47
 QY 105 KAEKPRRLATTASSAFEN-----TPFAASVLYQMPALNKGWLTPLKP 150
 DB 48 NISKFTKRALNTAAMQKRREHTVIDGNANPYKRAIAVVRKQLQRLDGEVYIKGNMP 107
 QY 151 LTPILSGA--ISGAMDQV-----GTQMDRARGDLHYLSTSPDKL--HDANAVSVKRR 199

Db 108 LKRSVLGAKAMATLHDKTVHVPFTRGKTRAQVLTALGQSLDEYGDGMDALATKIETP 167
 Qy 200 SPALGRQVDMGIAVQTFPSALNVVTVLAPALASRPSVOGAV---DFGVSTAGGLVANAG 236
 Db 168 GTVFQGI--KGLMFASQGLPIIAGLVPAIMAVANALGVVAGGALGVAGAFGAASGA 225
 Qy 257 FGRMLSVQSRDQLBGAFVLMGDKKEPKAALSE-ETDWDAYKAIKASVSGAALNAGK 315
 Db 226 FAFGAMAVSAIKMLNDGTLOATQATRRYQASLEQVKSTWEGIIKO-NOAQIFNTLSNA-- 282
 Qy 316 RMAGLPLVDATDGLKAVRSVLSATSLTKNGLALAGGVAGVSKLOKMATKNI-----TDS 369
 Db 283 -----LDTNVALGRMKPFL-----AGISKGHEQASQSVLKNAQNSQT 320
 Qy 370 ATK-----AAVSQSLNLSVGVFAG-----WTTAGLATDPVAKAESF 408
 Db 321 ASKPFNMNTTGVKTFNTLLSAGRFGDGLINVFQGLPLFWLTAAGL--DNLGKKFQNW 378
 Qy 409 I-----QKVKSTAST-----TSYVADQTVKLAKTV 435
 Db 379 ANSVAGQNAIKSFIEYTQTNLPKIGIFGNVPMGIGNLMKAFQAQNSNIFDNLVKMTAKF 438
 Qy 436 KDMGSAISSTG 447
 Db 439 REWSEQVKSEG 450

RESULT 10

US-09-085-199B-11
 ; Sequence 11, Application US/09085199B
 ; Patent No. 6235879
 ; GENERAL INFORMATION:
 ; APPLICANT: Hayden, Michael R.
 ; APPLICANT: Hackam, Abigail
 ; APPLICANT: Huq, A.H.M. Mahbubul
 ; APPLICANT: Chopra, Vikramjit Singh
 ; APPLICANT: Kalchman, Michael
 ; TITLE OF INVENTION: Apoptosis Modulators That Interact with the
 ; TITLE OF INVENTION: Huntington's Disease Gene
 ; NUMBER OF SEQUENCES: 44
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Oppedahl & Larson
 ; STREET: PO Box 5270
 ; CITY: Frisco
 ; STATE: CO
 ; COUNTRY: USA
 ; ZIP: 80443-5270
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS DOS 5.0
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/085.199B
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Larson, Marina T.
 ; REGISTRATION NUMBER: 32038
 ; REFERENCE/DOCKET NUMBER: UBC-P-01JUS2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (970) 668-2050
 ; TELEFAX: (970) 668-2052
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1068
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: no
 ; ORIGINAL SOURCE:
 ; ORGANISM: mouse
 ; FEATURE:

OTHER INFORMATION: Huntington-interacting protein -mHiPla
 US-09-085-199B-11
 Query Match 5.0%; Score 121.5; DB 4; Length 1068;
 Best Local Similarity 22.7%; Pred. No. 0.01;
 Matches 119; Conservative 74; Mismatches 178; Indels 153; Gaps 29;
 Qy 1 MHINQSAQQPPGVAMSGSFRFTASDASLASSVRSVSTTSCRDLOAITDYLKHHVFAHRFS 60
 Db 505 MKMEQSDQLEKLEKRLAARAGELARAEAL-SRTEQSGSELSRLDTLN-----AEKEA 558
 Qy 61 VIGSPDERDAALAH-----NEQIDALVETRNRLYSGETPATIAETFAKAEF----- 109
 Db 559 LSGVVRQREAEELAAQSLVREKEALSQEQRSQEQKELRGQLAEKESQEQGLRQKLKD 618
 Qy 110 DRLATTASSAFENTPFAAASVLMQMPAINKGDWLTATPLKPL-----TP--LISGALSG 161
 Db 619 EQLAVLRS-----AAAEAEAILQDAVSKLD-----DPLHLRCTSTSSPDYLVSR- 663
 Qy 162 AMQVGTGMDRARGDLHYLSTSPDKLHDAMAVSVKHSFALGRQVVDGMIAVQTFPS--- 218
 Db 664 ALDSVS-----GLEQGHQTYLASEDA--SALVAALTRFSLHAADTIYN-GAATSHLAPT 716
 Qy 219 -----ALNVV-----RTVLAPALAS--RPSVQGVADVRGVSTAGGLVANAG 256
 Db 717 PADRLMDTCECGARALELVGQLQDQTVLRRAPFSLMRAPLQGLQLGQDL----- 767
 Qy 257 FGDRMLSVQSRDQLRGAGFVLGMKDKPEKAAALSEETDMLDAYKAISA-----SYSGAA 310
 Db 768 ---KPKSLDVRQBELG-----AMVDKE--MAATSAAIE--DAVRRIEDMMSQARHESGVK 816
 Qy 311 LNAGKRWAGLPDVLDTGLKAVRSLV-SATSLTKNGLALAGGVAGVSKLOKMATKNI--T 367
 Db 817 LEVNERI-----LNSCTDLMKAIRLLVMTSTSQKE---IVESGRGAATQGEFVAKNSRWT 869
 Qy 368 D---SATKAAVSQSLNLSVGVGVAGVTTAGLATDPVAKKAESEFIQDKVK-----S 415
 Db 870 EGLISAKA-----VGMGATQL-----VESADKVLVLMHGKYEELIVCSHE 909
 Qy 416 TASSTTSYVADQTVK-----LAKTVKDMGSAISST 446
 Db 910 IASTAQVLAASKVKANKNSPHLSRLQECSTRTVNERAANVAVST 953

RESULT 11

US-08-441-139-2
 ; Sequence 2, Application US/08441139
 ; Patent No. 5773245
 ; GENERAL INFORMATION:
 ; APPLICANT: Wittrup, Dr. Karl D.
 ; APPLICANT: Robinson, Anne S.
 ; TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
 ; TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
 ; STREET: 400 Garden City Plaza
 ; CITY: Garden City
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 11530
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/441,139
 ; FILING DATE: 15-MAY-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/089,997
 ; FILING DATE: 06-JUL-1993

```

1  ATTORNEY/AGENT INFORMATION:
2  NAME: DigiGlo, Frank S.
3  REGISTRATION NUMBER: 31,346
4  REFERENCE/DOCKET NUMBER: 86444
5  TELECOMMUNICATION INFORMATION:
6  TELEPHONE: 516-742-4343
7  TELEFAX: 516-742-4366
8  TELEX: 230 901 SANS UR
9  INFORMATION FOR SEQ ID NO: 2:
10 SEQUENCE CHARACTERISTICS:
11     LENGTH: 682 amino acids
12     TYPE: amino acid
13     TOPOLOGY: linear
14     MOLECULE TYPE: protein
15     OS-08-441-139-2

```

Query Match	4.9%	Score 117	DB 1	Length 682
Best Local Similarity	21.2%	Pred. No. 0.014		
Matches 114	Conservative 64	Mismatches 165	Indels 194	Gaps 28

Qy	71	ALANHEQIDALVERFARNLVEGETPAIAT	FAKKEFEDRLATTASSAFE	131			
Db	65	AVMKNGKTEILANEGNRI	TPSVAVTDEBERLIGDAAKQVAMNPNTIFDIKRL	119			
Qy	122	-----NPFPAALSTQWOPAIN	KDMLATLPKLPTR LISGALSGA	162			
Db	120	IGLKRYNDSVOKIKHLPFNV	VNKOGKPAVEVSIVGE	KKFTPEELISGMILGK	172		
Qy	163	MDQV-----GTRMMDRARGDHLTSLTSPDKHDMMAVS	VKXHSBALGRQVVDMDIAVOTF	217			
Db	173	MKQIAEDYLGTRV	THAVVTYPAHYNDQROATK	DAG	210		
Qy	218	SALNVRTVLAPALASRPSVQGAVD	FGVSTAGS	LVANAG	FGDRMLSVQS	R	267
Db	211	AGLVNLRVNEPTA	ALAYGLDKDKEHOIIVYD	LGGTFEDVSLSTIENSVFR	263		
Qy	268	DOLRGAFVLMKMD		KEPKALSEE	291		
Db	264	VQATSGDTHLGEEDFPDYKIYNOQIKAPKKGIGIDVSDNNKALATL	KREAEAKARALSSQM	323			
Qy	292	-----TDWDAVYKAIKSASYSGAALNAGKEMAGI	PLDVAITDGLKAVRSIVSATSLT	342			
Db	324	STRIEIDSFVVGIDILSETLTRAKE	ELANDLDFPKTKLPKVEKVLQD8GLE	372			
Qy	343	K--NGLALAGYAVGSUO	KMATKNIT	DSATPAVASQLSNLVGSYGV	387		
Db	373	KKVYDDIYLVGSGSTRIPKVOQLLESYFDEGKASXGINDBEAVAYGA	VQAGVLSGEEVBE	432			
Qy	388	-----FAGWTTAGLATDPVAKKASEFI	QOKVSTASSTYSYAADQTVLAKLVKD	437			
Db	433	DIYLVDAVALTLGIETTGCVWTLPIKR	NTAIPIRK	KSQIPSTA	VDNQPTWIKIYXE	487	
Qy	438	MSGERAISTGASLASTVYNL		RHSAPBAD	EEGGSIAFSRSE	479	
Db	488	-----GGRAMSKONNLLGKTELTGIP	PAPAPGVQDIEETPALDNGILNKVATSD	535			

```

RESULT 12
US-08-809-326A-1
; Sequence 1, Application US/08809326A
; Patent No. 6165478
; GENERAL INFORMATION:
; APPLICANT: Izutsu, Hiroshi
; APPLICANT: Obara, Kazuhiko
; APPLICANT: Matsumoto, Akira
; TITLE OF INVENTION: CHLAMYDIA PNEUMONIAE ANTIGENIC POLYPEPTIDES, DNAS CODING
; TITLE OF INVENTION: FUSED PROTEINS CONTAINING THE POLYPEPTIDES, DNAS CODING
; TITLE OF INVENTION: THEREFOR, RECOMBINANT VECTORS CARRYING THE DNAS, TRANSFORMANTS
; TITLE OF INVENTION: CONTAINING THE RECOMBINANT VECTORS, METHOD FOR PRODUCTION OF
; TITLE OF INVENTION: ANTIBODY, METHOD AND REAGENTS FOR DETECTION AND/OR MEASUREMENT
; TITLE OF INVENTION: OF ANTIBODY, METHOD AND AGENTS FOR DIAGNOSIS OF CHLAMYDIA
; TITLE OF INVENTION: PNEUMONIAE INFECTIONS, PROBES AND PRIMERS FOR DETECTION AND/OR
; TITLE OF INVENTION: MEASUREMENT OF CHLAMYDIA PNEUMONIAE GENE, AND METHOD AND

```

TITLE OF INVENTION: REAGENTS FOR DETECTION AND/OR MEASUREMENT OF CHLAMYDIA
 TITLE OF INVENTION: PNEUMONIAE GENE
 NUMBER OF SEQUENCES: 31
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/809,326A
 FILING DATE: 19-MAR-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 224711/94
 FILING DATE: 20-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 106006/95
 FILING DATE: 28-APR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 106008/95
 FILING DATE: 28-APR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 106010/95
 FILING DATE: 28-APR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 106009/95
 FILING DATE: 28-APR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 106010/95
 FILING DATE: 28-APR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 106011/95
 FILING DATE: 28-APR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Miller, Charles E.
 REGISTRATION NUMBER: 24,576
 REFERENCE/DOCKET NUMBER: 7426-043-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 488 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 JS-08-809-326A-1

	Query Match	4.8%;	Score 116;	DB 4;	Length 488;	
	Best Local Similarity	21.6%;	Pred. No. 0.0099;			
	Matches	95;	Conservative	59;	Mismatches 157;	Indels 128; Gaps
QY	63	GSPDERDALAHNEQIDALVETRANRLYSEGETPATIAETFAAEKEFDRLATTASSAFEN	1222			
Db	26	GVFPQODDLKSGNETKQID---QTRQSK-NTEMSDATIAGSGDK-----TSSTTKTET	75			
QY	123	TP---FAAASVLQYMGQA-INKKDMLATPLKPLPLISGLASGAMQVGRKM-----	170*			
Db	76	APQGVAAAGSESSSQCGADTG-----VSGAAATTTASNTATKIMQSIIE	1222			
QY	171	-----MDRARGDLHLSTSPDKLHDMAAV-----SVKHSHPALGR-----QVDM	210			
Db	123	ASKSMESTLESIGLSIAAQKKEVBAVVVALSGKSSGSAKLETPELPKPGVTPRSEVIEI	1822			
QY	211	GIAVQTSALANVVRTVLAPLALASRPSVQGVADVGVSTAGLVANAGFGDMLSV-QSRDQ	2659			
Db	183	GLAAL--KAIQTQGEATKSLASVNSTAOAD-----QTNKLGLEKAIKIDIEREE	2322			


```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,467B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hereko, Bart S.
; REGISTRATION NUMBER: 32,572
; REFERENCE/DOCKET NUMBER: 5474R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-0633
; TELEFAX: (513) 627-0260
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2887 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-462-467B-8

```

```

Query Match 4.8%; Score 116; DB 4; Length 2887;
Best Local Similarity 21.0%; Pred. No. 0.19;
Matches 117; Conservative 76; Mismatches 217; Indels 146; Gaps 22;

```

```

QY 24 ASLAASSSVRSV-TTSCRDLOAITDYKHHVFAA--HREFSVIGSPDERDAALAHNEQID 79
DB 1027 ASPASNNALALILESERGVVALGLY-THRILEARGYRMTALAP-RGVALLLEGGLY 1084
QY 80 ALVETRANRLYSEGETP--ATIAETFAKAEKFDRLATTASAPENTFPAA---SVLQYM 134
DB 1085 ALAVALLASNLNLEAGASGCSYGLSERLALALBELYGLNVALSPMETTYRALBELGLYTEIL 1144
QY 135 QPAINKGDMLATPLKPLTPIISGALSGAMDQVGTOMDRARGDLHYLSTSPDKLHDMAV 194
DB 1145 ETYRTREGVALAHE-----METARGCYSTHRASPLEPHEBERGL 1183
QY 195 SVRHSHPALGRQVDMGIAVQTFSSALNVRTVLAAPALASRPVQGAADFVGVSTAGGLVA- 253
DB 1184 YGLSERVALPR-----ASPTYRGLN-METALAP-----HEGLNTHRGVAL 1223
QY 254 -----NAGRGDMLSVQSRDQLRGAFVUGMDKEX----- 285
DB 1224 GLYASNHISPRTRHPHEGLASPMETGVALLEVALSERARGGILYGLNARGERLYSPH 1283
QY 286 -----AALSEETDMLDAYKAIKASAYSQAA----- 310
DB 1284 EPPGLALATRPYSGLASNSERLEALAVALARSGERLEYGLTHRILEGLASCYSTRP 1343
QY 311 ----LNAGKMAGLPLDVAITDGLKAVSLVSATSLTK-NGIALAGG---YAGVSKLQKM 361
DB 1344 ASPGLNASPALAGLAL--AARGLETHRALAGLNCYSALAGLARGMETALAGLLEMETM 1401
QY 362 ATKRIITDSATKAIVSQ-LSNLVGSVGVFAG---WTTAGLATDPAVKKASFIODKXST 416
DB 1402 ET--ILE--TRPGLARGASNLYSERVALSERPTRHVALASNRMETISBRTTHALMET 1457
QY 417 ASSTTVADQVTKLAKTVDMGSEALISGTASLRSTVNN-----BHRSAPEADIE 468
DB 1458 GLNASNGLAGASNLBERERIASNMGARGAVALLPRLYSILEGLYPTTYTPRPSPTYRSE 1517
QY 469 EGGISAFSRSETPQL 484
DB 1518 RSERS--ERSERTYRI 1531

```

```

RESULT 15
US-08-591-079-6
; Sequence 6, Application US/08591079
; Patent No. 5972899
; GENERAL INFORMATION:
; APPLICANT: Zychlinky, Arturo
; APPLICANT: Chen, Yajing
; TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab

```

```

; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FORSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,079
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lynam, Samuel
; REGISTRATION NUMBER: 33,949
; REFERENCE/DOCKET NUMBER: 15661-20017.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0764
; TELEX: 90-4030 MRSNPOERSWSH
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-591-079-6

```

```

Query Match 4.8%; Score 115.5; DB 2; Length 401;
Best Local Similarity 21.2%; Pred. No. 0.0079;
Matches 88; Conservative 51; Mismatches 120; Indels 157; Gaps 21;

```

```

QY 122 NTFPAASVLYQYQ---PAINKGDMLATPLK-----PLTFPLISGALSGAM 163
DB 10 STP-VTGSLLPVEYETPAPAPLQVQVAGELKDNKGVSQGVQLPAPLAVVASQVTEGQQ 68
QY 164 DQVGYKMDR-ARGLD-----HYLS-----TSPD----- 186
DB 69 QEV-TKLLESVTRGAAGSOLISNYVSLTKFTLASPDTEIEIBGLVSNLEVRKDIKIA 127
QY 187 ---KLDMAVSVKRSRSPAL-----GRQVDMGIAVQTFSSALNVRTVLAAPALASRPV 237
DB 128 DIQRLHQNMKKLENOEKIKETEENAKOVKSGIASKIFGWLASIASVIGAIMVASGV 187
QY 238 QGAVDGVSTAGGLVANMAGFDRMLSVQSRDQLRGAFVUGMDKEPKKALSETDMLDA 297
DB 188 -GAV-----AGAMMYASG-----VIGM-----A 204
QY 298 YKAIKASYSGALLNAGKMAGLPLDVAITDGLKAVSLVSATSLTKNGIAL-----AGGY 352
DB 205 NMAVKQAAEDGLISQAMKILG-PILTAIE---VALTVSVTWTFGGSGALKCIANIGAK 259
QY 353 AGVSKLQKMAKRIITTSATKAIVSQ-LSNLVGSVGVFAGTTTGATDPAVKKASFIOD 411
DB 260 LG-ANTASLAAGAERSAKVAQISTGISMVGSAAVTKLGSGSFAGLTMSHAIR----- 310
QY 412 KVKSTASSTSVADQVTKLAKTVDMGSEALISGTASLRSTVNN-----LRHSA 462
DB 311 ----TGSQAT-----QVAVGVSIGITGITINNKQADLOJHNN 343

```

```

Search completed: January 2, 2003, 15:21:49
Job time : 21.5139 secs

```

